IMPROVING TARGET ORIENTATION DISCRIMINATION PERFORMANCE IN AIR-TO-AIR FLIGHT SIMULATION

by

Gary Serfoss

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

May 2003

*The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
29052003	N/A	-

Title and Subtitle Improving Target Orientation Discrimination Performance in Air-to-air Flight Simulation	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Arizona State University	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 77	

Chapter I

Introduction

Overview

When performing air-to-air combat within visual range of the target aircraft, pilots are in an extremely high workload environment where they are pushing their own limits, as well as that of their aircraft. Nearly all attention is focused visually on trying to acquire, track and assess the adversary's orientation and range. If a pilot detects the enemy just a second or two later than the enemy detects them, or misjudges the orientation of the enemy aircraft and falsely concludes how he is maneuvering - the enemy will likely shoot him down first.

The problem facing modern flight simulators when they try to simulate this scenario is that, despite recent gains in fidelity, they are unable to provide the necessary resolution to accurately represent enemy aircraft at realistic ranges. For example, the fine details of an aircraft seen flying in the real world typically are not visible in the simulator at the same simulated distances. Pilots use these details to help determine things like the pitch, bank and heading of the target aircraft. It can be argued that for pilots to receive realistic training on how to tactically maneuver in these situations, they need to have a simulated enemy that gives them real-world information at real-world ranges. Not having this information until the enemy is much closer can change how the pilot maneuvers and almost certainly affects the outcome of that maneuver compared to what it would have been had the enemy's orientation been better assessed at a more realistic distance. For example, the best method for getting in behind an enemy as he approaches at 4 miles, will likely be much different than trying to accomplish the same maneuver

when the enemy is only 1 mile away. This is especially true with the tremendous closure rates experienced with modern jet aircraft. If, due to lack of resolution in the simulator, the pilot is having to artificially wait much later to engage the enemy and having to modify how he engages him, then proper training likely is not taking place. In fact, there may be a negative transfer of training as bad techniques learned in the simulator are later transferred to actual flying. Given this problem, the difficulty then becomes trying to reach some type of solution with the simulator that is visually acceptable, yet also practical from a cost and technology standpoint. How can the simulator be made to more accurately depict what a pilot sees in a real-world environment?

The Solution. Somehow the visual projection systems used to present target aircraft need to be improved or the targets themselves need to be modified so that detail can be discriminated at more realistic distances. Some of the smaller details of the plane visible in real-world conditions, can not be seen in the simulator. Pilots use these details to help determine what the target aircraft is doing and then react appropriately. As will be discussed later, better projectors, or specialized projectors that only display a high-definition target, were not viewed as viable solutions at the time of this study. Therefore, with no simple or cost-effective means of improving the projection system, modifying the targets in the current system seemed to be the most feasible option. The solution chosen here involves artificially enlarging the size of the simulated target to make the necessary cues more visible. It was believed that the target could be enlarged enough to do this, yet not be so large as to appear unrealistic or give the pilot a false sense of how the target would look in real combat. More details of this proposed solution are discussed in Chapter III. First, however, is a discussion of the reasons that better or specialized

"target" projectors were not seen as immediately viable options. This is followed by a background discussion of air combat in Chapter II. Chapter II also addresses what is known about detection distances in air combat, what cues pilots report using to judge aircraft orientation and at what distances those cues can be seen.

Better Projectors. An obvious solution to the lack of detail problem would be to simply use a display with better resolution. However, currently this wasn't a practical option because projection technology hasn't reached the necessary level yet. Current state-of-the-art systems still lack the level of detail needed and there is little promise of a technological solution to this problem within the next few years (Pierce, 2002). As of 2002, Barco's 808 projectors represented near-leading edge flight simulation display technology (Miller, 2002). However, they still only provided resolutions roughly equal to 20/40 visual acuity. This is more than adequate for training most flight tasks but fails to provide the air combat target information that a pilot can only gather at the threshold of their 20/20, or better, vision.

Another reason technology may not provide an immediate solution is the high cost of state-of-the-art equipment. Because of that high cost, even after the necessary technology exists, it will likely be several years before it is a viable option even for the larger, better-funded organizations that use simulators. Furthermore, the industry trend has been away from expensive, high fidelity systems, even by the U. S. Air Force. Therefore, fewer and fewer users are willing to purchase the ultra-expensive high-end systems. Packaging this technology so that it costs thousands instead of millions will take even longer. This argues the need for an interim solution (such as proposed here) that can be used over the next several years or more.

Target Projectors. Another promising solution for adding display detail is the use of target projectors. These specialized projectors have been developed to provide a high-resolution outline of a target aircraft on the viewing screen of the simulator. However, they are extremely costly and also have limitations that keep them from being a complete solution. The concept of target projectors is to allow the visual scene to be displayed using a conventional projector then use a specialized projector to superimpose the image of a target aircraft against that scene. Since all the target projector has to display is the target aircraft, it can be built specifically for that purpose and the result is a very sharp, high-resolution outline of the target aircraft. Nevertheless, as mentioned, they have several shortcomings. First, is the problem of cost. A high quality simulator display without a target projector may cost $500,000. The same display with a target projector would cost closer to $1.2 million (Miller, 2002). Furthermore, purchase price is not the only cost issue. Target projectors are also very expensive to operate and very hard to maintain, which significantly raises operating costs and reduces system reliability. Secondly, their performance is limited. Currently they can only project a monochrome image that appears artificial against the simulated flight environment. This causes artificial cueing, where the pilot's eye is drawn to it because it stands out from the rest of the simulated scene. Real-world targets usually blend into the visual scene much better. Therefore, artificial cueing largely defeats the goal of trying to replicate actual air combat. Another problem is that air combat often involves multiple aircraft (both friendly and non-friendly) and target projectors can only present a few. When a target projector is being used to project a target, remaining targets are displayed using only the existing visual scene projector. Therefore, as the target projector transitions from

showing one target to showing another, the affected target aircraft very noticeably pop in and out of view or change dramatically in appearance. Again, the effects are so noticeable that it negatively impacts the training the pilot is receiving (Miller, 2002). A final problem is that the target projectors only provide a sharp outline of the aircraft. This is fine for target cues that are recognized as variations in the aircraft's outline, like wings, nose, tail etc. However, other cues are often seen within the outline of the aircraft. They are visible due to their contrast with other parts of the aircraft body, not because they significantly affect the overall outline of the aircraft. Some examples are the canopy, intakes, exhaust, missiles and so on (Warner, Serfoss, & Hubbard, 1993). At times, these cues are silhouetted against the sky and affect the outline, but most times they are seen against the background of the plane and yet, are still distinguishable. Target projectors lack the ability to show these types of cues.

Best Remaining Solution. If a better visual system is not available/practical, and a supplemental system like target projectors is not a good solution, the best remaining solution seems to involve enhancing or modifying the current target aircraft model in some way to achieve the desired results. How could this be done? There may be dozens of solutions. However, one appears to be somehow enlarging the target aircraft so that it and its' cues can be more easily seen. Other options that may help include only enlarging portions of the target that are known to be needed as cues (instead of the whole aircraft) or perhaps enhancing the needed cues using color, shading, or some other variable. These last two approaches have some limitations that size enhancement doesn't appear to have – they will all be discussed more in Chapter III. Nevertheless, finding an effective,

yet subtle and believable way of artificially enlarging the target aircraft seems to be a good starting point.

The goal of this study was to investigate the feasibility of using **increased target size** to provide the necessary cues for more accurate air combat simulation. If increasing target size is only partially successful, the use of other enhancements such as color can possibly be used at a later date to further augment the model. The goal is that if a pilot typically used the presence or absence of a cue (canopy, engine intake, etc.) when an aircraft is at a certain distance and angle to determine its orientation, that some modification can be made to allow that cue to be seen, if present. As mentioned, the details of how the target size is increased are discussed in Chapter III.

Purpose and Scope

The primary objective of this investigation is to provide better target cueing for close-range, air-to-air combat engagement simulations by increasing target size. The goal is to determine how much of an increase in target size is necessary to reach real-world performance goals as illustrated in Chapter II. For example, size should be increased enough to be able to make accurate orientation judgements at the appropriate ranges as discussed in Warner et al. (1993).

The results gathered here will help determine what size an aircraft needs to be to provide quality cueing for air-to-air combat in current high-fidelity simulators. An improved depiction of the target will remove several of the current limitations of air combat simulation and allow better, more realistic training to pilots. It should also apply to other fields of simulation where fine target detail is needed but is not readily achievable, ranging from medical simulations to video games. Finally, the benefit of this

type of work could also be readily applied even when some systems do exist that *are* capable of giving the necessary level of resolution and detail, since that ability will likely only come at a great monetary expense. If this type of manipulation can allow the use of less expensive equipment, then it can literally save users millions of dollars.

Chapter II

Background

Air Combat Simulation

Flight simulation is a critical resource for training pilots in air combat for several reasons. First, due primarily to budget constraints, pilots often have a limited number of flying hours each month and therefore, limited opportunities to train skills like air combat maneuvering. Secondly, close-range (less than 5 nmi) air combat can be very dangerous. Planes and pilots are flown at their limits, sometimes causing unconsciousness, aircraft failure or the pilot placing the aircraft in an unrecoverable flight condition. Mid-air collisions are also a real threat, as are collisions with the terrain, since nearly 100% of the pilot's mental resources may be focused on the enemy in a kill-or-be-killed environment. Finally, actual air combat flight training happens so quickly and in such a complicated way (especially with four or more aircraft involved) that it is very difficult to recall all the necessary details during a debrief of the mission. This lack of detail and accurate visualization can limit how much and how quickly a student can learn from a given engagement.

Flight simulators help overcome all of these concerns. They cost much less than an aircraft to operate, and also allow the pilots to train on only the skills they want at the time. They can be reset again and again to quickly allow the pilots to try a tactic again or engage in a new scenario. Simulated training also provides a safe, risk-free environment where pilots can be allowed to make mistakes and see their consequences without putting themselves, or other pilots or aircraft at risk. Finally, the entire mission can be saved and replayed in complete detail, allowing a student to see exactly what they and others did.

They can view the engagement from several different perspectives, and can even see a replay of how an expert would fly a similar scenario. A further benefit is being able to pause the action at any point to allow an instructor to give immediate feedback/correction when it can most easily be illustrated and remembered. Certainly there are several other benefits to flight simulation, but these should be the most obvious relative to air combat.

Known Target Detection Capabilities

Several investigations have been made to determine aircraft target acquisition capabilities. A total of 759 training engagements at the Naval Air Station Oceana Tactical Air Combat Training System (TACTS) range revealed that in 624 of the engagements the pilots first sighted the target as a dot against the background at an average distance of 5.67 nmi (Hamilton & Monaco, 1986; Monaco & Hamilton, 1985). In the remaining 135 engagements exhaust smoke, contrails and sun glint off the aircraft allowed the pilots to detect the aircraft at even greater distances. In the 122 engagements where exhaust smoke was the primary cue, detection distances averaged 7.64 nmi. Environmental and local conditions as well as target type and paint scheme play a significant role in detection distances though. Variables such as background sky or ground coloring vs. aircraft coloring, brightness and directness of sunlight as well as target location vs. the sun and several other variables can either enhance or decrease detection distances. Furthermore, although Hamilton & Monaco found several instances where exhaust smoke was the primary cue, this condition is arguably becoming of decreasing value as aircraft emissions have become less visible over the last decade or two. Table 1 provides a large list of factors that have been shown to affect target detectability. These items were taken from the field evaluations cited in this section as

well as Bloomfield & Smith (1982), Boff & Lincoln (1998), Buffett (1986), Costanza, Stacey, & Snyder (1980), and Hoffman (1976).

In 1983, Kress & Brictson studied 87 air-to-air engagements at the Yuma TACTS range. Average unaided detection distances for the target F-5 and F-4 aircraft were 3.1 nmi. When the pilots were aided with a head-up display (HUD) symbol that cued the pilot to the target's location, the mean detection distance grew to 6.8 nmi.

Another study that investigated detection distances was Temme & Still (1991). They measured air-to-air target detection distances at the Naval Air Station Oceana TACTS range to see if there was a performance difference between those pilots who wore corrective eyeglasses and those who did not. Those with eyeglasses did not detect the targets until they were about 10% closer than those with unaided vision. Two very closely matched groups of eyeglass and non-eyeglass wearers had average detection ranges of 4.52 and 5.64 nmi respectively when using all detection means including aircraft sighting, target glint, contrails and exhaust smoke. When limiting subjects to aircraft-only detections, the corresponding distances were 4.35 and 5.54 nmi respectively. Although the distinction of glasses vs. no glasses is not of interest to this investigation, it does provide two more data points for detection distance ability.

Another study by Hutchins in 1978 at the Air Combat Maneuvering Range (ACMR), which is the earlier name of the TACTS, involved 45 air combat training engagements. The mean detection distance of the A-4 targets was 3.09, with a range of 0.38 to 6.23 nmi. Other studies were done using observers on the ground. With visibility conditions spanning 7 to 10 miles over an 8-day testing period, O'Neal & Miller (1998) found detection distances for approaching T-38 aircraft to ranged from 4.77 to 6.73 nmi.

Table 1

Factors Governing Target Detectability

<u>Target Features</u>	
Size	Motion
Shape	Exposure time
Brightness	Smoke
Coloring	Contrails
Type of aircraft	Glint
Location in the visual field	Dynamics
Target-to-background contrast	
<u>Environmental Characteristics</u>	
Background brightness	Cloud cover
Sun position	Glare
Background clutter	Cockpit structures
Atmospheric conditions, i.e., haze, smoke, dust	
<u>Observer Characteristics</u>	
Visual acuity	Workload
Accommodation and myopia	Eye movements
Contrast sensitivity	Training
Search/scan patterns	Contrast sensitivity
Miscellaneous factors, i.e., motivation, fatigue, anxiety, anticipation	Individual differences
<u>Observer Aircraft</u>	
Dynamics	
Cockpit physical structures	
<u>Target and Observer Aircraft Relationships</u>	
Slant range	
Target orientation	

Note. From "Visual Cue Requirements for Target Orientation Assessment in Air Combat Simulation" by H. D. Warner, G. L. Serfoss, and D. C. Hubbard, 1993, AL-TR-1993-0021, Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory, p 5. Adapted with Permission.

Another ground observer study used 400 visual detections of a T-38 aircraft (Provines, Rahe, Block, Pena, & Tredici, 1983). The aircraft was approaching from a known direction and a distance of 9 miles and mean detection distance was 4.55 miles over the 400 trials.

A final note about detection distances is that actual detection distances for target aircraft have been found to be considerably less than would be predicted theoretically. For example, the previously mentioned Hamilton & Monaco (1986) and Monaco & Hamilton (1985) studies found that the exposed amount aircraft needed for detection was about four times larger than mathematically predicted based on the subjects' performance on two vision tests for high contrast acuity and visual detection thresholds. Several environmental, vision and flight performance factors were believed to account for this disconnect.

A similar phenomenon has also been observed in flight simulators where actual performance is often significantly lower than would be predicted theoretically. Some of this has to do with the physical limitations of the projection systems and the target only being large enough to be presented using between 0-2 lines of resolution. The presence of the target in each line of resolution is not gradual, the pixel(s) are either 100% on or 100% off. When the target is presented dynamically moving from line to line of the display near the threshold of visibility it is common for aliasing to occur. Aliasing is basically where an object, like the target, seems to pop on and off of the display screen. One second it is there and the next it is not. It is near sub-pixel size and since it is often being "painted" on the screen by only a line or two of video, it is really at the projector's threshold, thus causing it to keep popping in and out of the scene.

This same principle also applies to other smaller visual cues on the target as it gets closer. For example, although the target may be easily seen at two miles, the pilots are often looking for the location/presence of the canopy, engine intakes and other features. As these smaller features are at the threshold of visibility they also experience the effects of aliasing. Since with current technology this rough depiction of target features at threshold detection ranges cannot be readily changed, it is a goal of this study to design a model whose mean performance, in spite of aliasing, more closely resembles the real-world performance that a pilot would expect.

As mentioned above, detection is not only important for detecting the aircraft as a whole, but it is necessary in identifying the presence/absence of certain aircraft features that allow a pilot to determine the orientation and possible intended actions of the enemy aircraft. Since the aircraft is already detectable at realistic distances, what was primarily of interest here was whether the smaller cues of the aircraft used for discriminating its orientation were present and visible at the appropriate ranges. A discussion of the useful aircraft features and the performance they provide at various distances follows.

Target Orientation Assessment Capabilities and Cues Used

As alluded to above, target detection is only a small part of air-to-air combat. Once the target is detected, the true art of air combat “dog fighting” begins. It then becomes a chess match to the death. Each pilot’s aircraft has different strengths and weakness, and different countries and groups of pilots often employ different tactics for conducting an aerial dogfight. However, other than equipment, training and talent/intelligence, success really depends on being able to determine what the enemy is doing, in order to help anticipate what he plans to do next, and thus attacking or reacting

appropriately. Discerning what the enemy is doing requires visually judging the orientation, speed and altitude of his aircraft. Orientation discriminations are often the most difficult of these tasks. Pilots usually do not have 10 seconds to watch the enemy before reacting - they need to react within a couple of seconds or less. Therefore, they need to quickly determine many things: is he heading left or right?; up or down?; toward them or away?; flying level or turning?, and much more. Of course, to accurately depict the important aspects of the enemy aircraft it is necessary to determine which features of the target, pilots are using to answer these questions.

Coward and Rupp (1982) gathered input from 15 F-4 fighter pilots to see what visual cues they used in real-world air combat (as estimated by their actual air combat training flights). Pilots cited relative motion on the canopy, wing planform, target nose position and relative size/size changes as the most important cues – rating them as "vital" more than 50% of the time. Warner et al. (1993) expanded on this by evaluating what cues were used as a function of aircraft orientation. They used high-resolution slides projected on a screen for pilots to evaluate. They found the most important cues were the tail, wings, nose, engine intake(s), planform, canopy, belly, missiles, top of aircraft, fuselage, and exhaust outlets. Of course, which cues were most important depended on which orientation discriminations needed to be made. For example, when the aircraft was pointing straight at, or straight away from the subject, cues like missiles, planform, top vs. belly etc. did.

Discussion of Metrics. Before continuing with a discussion of cues, it will be beneficial to discuss some of the measurement variables typically used to describe an aircraft's orientation. They can be condensed to three basic descriptors: pitch, bank and

aspect angle (although aspect angle often is further split into left vs. right). Pitch and bank involve the orientation of the target, with the ground or a stationary subject as the reference point. Pitch can have varying degrees, but for the purpose of this investigation is simplified to: up, down and level. It refers to the orientation of the nose of the aircraft vs. the ground (or reference of level). An aircraft flying with it's nose parallel to the ground is at a level pitch; if it's nose is pointed up away from the ground, it is pitched up; and if its nose is pointed down toward the ground, it is pitched down. Bank is similar, but instead of comparing the longitudinal axis of the aircraft to the ground it refers to the aircraft's rotation about that longitudinal axis vs. the ground. For example, when the aircraft is not tilted left or right (from the perspective of a pilot flying the aircraft) and both wings are equi- distant to the ground, there is no bank. When the aircraft is tilted to the left and the left wing is nearer the ground than the right, the plane is in a left bank and vice versa for right bank. Finally, aspect angle refers to the angle formed by the target flight path and the line-of-sight from the subject to the target when measured from the tail of the aircraft (Department of the Air Force, 1996; Murray, 1987). When the subject is stationary, as in this investigation, the aspect angle is fairly easily shown as a type of compass heading as shown in Figure 1. Armed with these definitions, one can now better understand some other findings from Warner et al (1993).

The Warner et al. (1993) study not only showed the cues the pilots used, but as importantly, showed how often the pilots correctly identified the pitch, bank and aspect angle of the target for each of 16 different orientations used in the study. Table 2 illustrates some of these findings, both for the F-15 and F-16 targets. The numbers shown in this table are the percentage of 40 subjects who got the orientation completely

correct (pitch, bank, and aspect); it uses a right vs. wrong scoring method instead of the 0-7 scale used in this study.

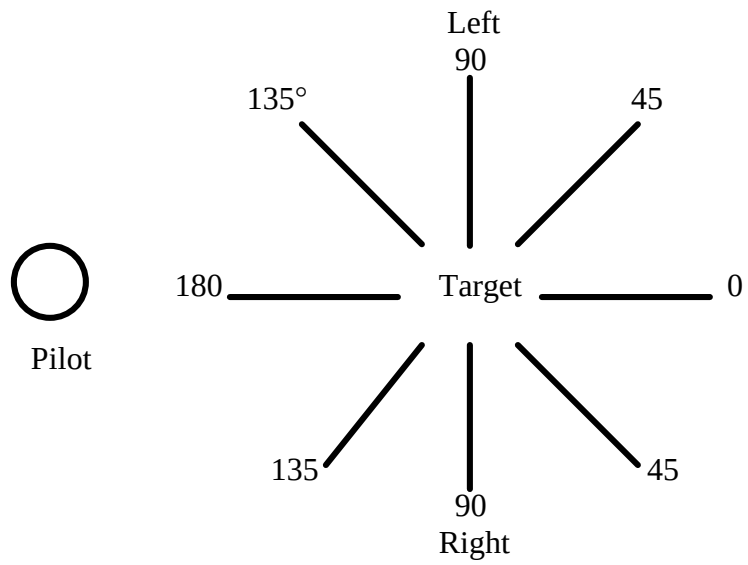

Figure 1. Illustration of how aspect angle/heading information were defined for this investigation and Warner et al. (1993).

---

[1]From "Visual Cue Requirements for Target Orientation Assessment in Air Combat Simulation" by H. D. Warner, G. L. Serfoss, and D. C. Hubbard, 1993, AL-TR-1993-0021, Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory, p 12. Reprinted with Permission.

These measures are considered critical to this investigation since the primary goal was to provide a good simulation of real-world air combat, and that is impossible to judge without valid performance metrics with which to compare. As discussed in Chapter III, this investigation was conducted to try to determine what size the target had to be at these four distances to achieve performance equivalent to that shown in Table 2, which is data from the Warner et al. study where subjects viewed high resolution slides.

Table 2

Percent Correct Responses as a Function of Target Orientation, Type and Distance

	F-15				F-16				
	Target Distance (nmi)				Target Distance (nmi)				
Target Orientation	0.5	1.0	2.0	3.0	0.5	1.0	2.0	3.0	Mean
6	95.0	97.5	95.0	82.5	95.0	80.0	80.0	67.5	86.56
9	95.0	92.5	87.5	87.5	80.0	77.5	67.5	67.5	81.88
2	95	95	97.5	87.5	57.5	60.0	77.5	72.5	80.31
5	95	97.5	95.0	85.0	90.0	72.5	52.5	45.0	79.06
13	100.0	95.0	72.5	60.0	90.0	85.0	50.0	15.0	70.94
12	92.5	92.5	77.5	65.0	72.5	42.5	50.0	57.5	68.75
3	95.0	82.5	62.5	52.5	90.0	57.5	42.5	42.5	65.63
7	97.5	95.0	45.0	25.0	90.0	82.5	47.5	25.0	63.44
10	77.5	70.0	62.5	40.0	80.0	70.0	32.5	32.5	58.13
4	97.5	92.5	30.0	37.5	90.0	67.5	27.5	15.0	57.19
1	47.5	35.0	40.0	45.0	65.0	52.5	37.5	35.0	44.69
15	97.5	62.5	32.5	27.5	62.5	27.5	20.0	10.0	42.50
8	75.0	67.5	20.0	17.5	32.5	25.0	30.0	5.0	34.06
14	72.5	37.5	27.5	15.0	17.5	10.0	20.0	17.5	27.19
16	65.0	55.0	20.0	12.5	30.0	7.5	12.5	2.5	25.63
11	60.0	40.0	5.0	17.5	40.0	17.5	5.0	0	23.13
Mean	84.84	75.47	54.38	47.34	67.66	52.19	40.78	31.88	56.82

Note. From “Visual Cue Requirements for Target Orientation Assessment in Air Combat Simulation” by H. D. Warner, G. L. Serfoss, and D. C. Hubbard, 1993, AL-TR-1993-0021, Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory, p 34. Reprinted with Permission.

Current Simulation Capabilities/Limitations

Several hardware/software limitations were mentioned in Chapter I and are also addressed in Chapter III where details of the proposed solution are discussed. However, another important capability/limitation distinction needs to be made here. Although lack of target detail could conceivably be due to the software model itself lacking sufficient detail, this is decidedly not the case here - the limitation is clearly due to the video projector's limitations. One can inspect the software model of the plane at close range and see that it is indeed very detailed and looks very much like a real plane. It is also programmed to remain detailed at all distances. Nearly every curve and contour of the real aircraft is present. Therefore, the problem is not that the necessary detail is not present on the model, but rather that the projector cannot show nearly all the detail when the object is small.

Past Solution Attempts

In the past, model sizes have been increased to 150% and more of actual size to see if this would make the necessary cues more visible and the aircraft easier to detect. There was some small benefit at the greater distances, but this was more than outweighed by how unrealistically large the model appeared at closer ranges. The reason that a 150% change is not more noticeable at threshold detection distances will be explained in more detail in Chapter III. However, one must keep in mind that when an object is so small that it covers only a few lines of video, it takes significant increases in size just to get coverage from another line or two of video. This is why the constant size increase approach (where an aircraft is multiplied by the same amount (i.e. 150 percent) all distances) did not seem viable. Instead, this study used a non-constant sizing factor that

only increased size slightly at close ranges and increased size much more at a distance. This method is discussed in the following chapter.

Chapter III

Proposed Solution

As mentioned above, some informal attempts have been made in the past to increase the target aircraft size by some constant multiple without success – the target seemed much too large at close ranges and not large enough at a distance. Another method of highlighting cues could be the use of color or shading. However, when the target and/or associated cues are not even visually present at the desired distances, there is no point trying to color or otherwise highlight them. It is not that the target or cues are technically present but just not detectable by the pilot – they are physically not showing on the screen. Therefore, in order to fit the target or necessary cues on the screen, increasing their size seems to be the only solution. Certainly, once the targets are physically present, making sure they are detectable could involve other types of enhancements besides further size increases (i.e. color, shading etc.).

If size is to be used to make the targets and cues more visible, and using some constant size increase is not acceptable, the solution lies in altering size in a way that it **is** acceptable. It seems possible that if the target could be magnified only minimally at close ranges yet magnified significantly (perhaps 2X or more) at a 3-mile distance, the desired cue visibility could be achieved. This is illustrated in Figure 2. Figure 2 shows a hypothetical example of how, at close range, where the target and necessary cues are readily visible, the aircraft is only magnified slightly. However, at ranges of 3 miles and more, where cues may not be visible, it is magnified by 2X or more. This method of altering size allows the aircraft and its associated cues to be visible at more realistic distances. However, it is important to note that the aircraft is still decreasing continually

in size as it goes further away, it just does not decrease in size as quickly. This is illustrated in Figure 3. The curve in Figure 2 comes from assuming that the target aircraft may need to be magnified 1.5 times at a hypothetical distance of 6000 feet and doubled in size at 14000 feet. By fitting a curve to those points and through the origin one gets the curve seen in Figure 2. The magnification estimates and distances are not based on any precise data or formula. However, they come as a “best guess” using pilot inputs and observing how large the images actually are at the various distances in the simulator and how large they likely may need to be to be seen properly. In some ways, the goal of the approach was to gradually magnify the target as much as possible while still ensuring that it would still noticeably decrease in size as it got further away. It is not a simple matter of just mathematically determining the proper size based on visual angle because, due to aliasing and other anomalies in the simulator, measured size in the simulator does not always match what the math says it should be. Furthermore, much debate and effort goes into even trying to properly measure objects in the simulator, which will not be discussed here. Nevertheless, this magnification curve was the best guess at least for a starting. Consequently, the desire to conduct this study was to gain empirical evidence to support or refine this curve. The result would hopefully be a curve that would be closer to ideal and also more defensible to critics of the approach.

As mentioned, Figure 3 helps illustrate that the target still does get smaller as a function of distance. Using basic geometry one can calculate the “unfactored size” line. The “factored size” then represents the hypothetical factor used in Figure 2. This target does not get as small as quickly. One can see that at 6000 feet the factored model is roughly 1.5 times larger than the unfactored one and twice as large at 14000 feet, which

were the magnification anchors discussed above and shown in the box for Figures 2 and 3. Please also note that the angular size of the target is roughly halved each time the distance is doubled. For example, the unfactored target is about 10 mils at 5000 feet and 5 mils at 10000. This assumption was verified in the simulator and used as a basis of determining the final magnification curve using data from this investigation – discussed in the next few chapters.

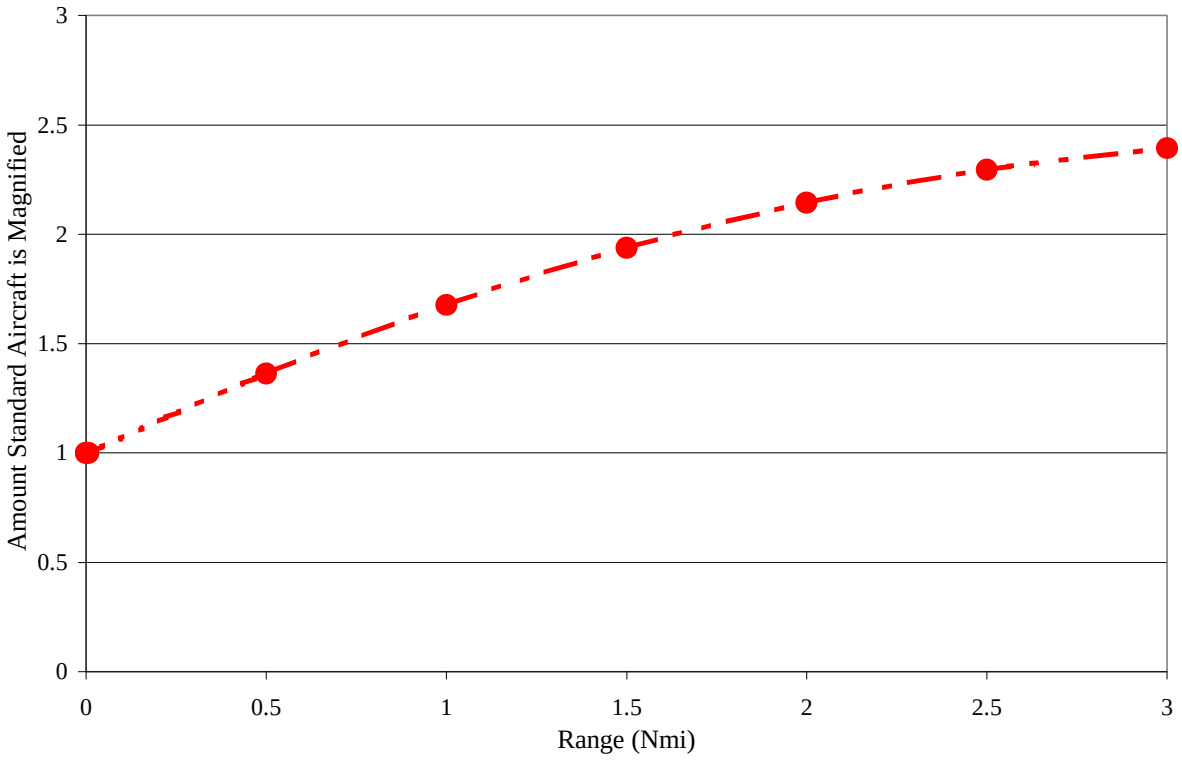

Figure 2. Example of possible target magnifications as a function of distance from observed

[2]From D. Lerman, Warfighter Training Research Division, Air Force Laboratory, Mesa, AZ. Reprinted with permission of the author.

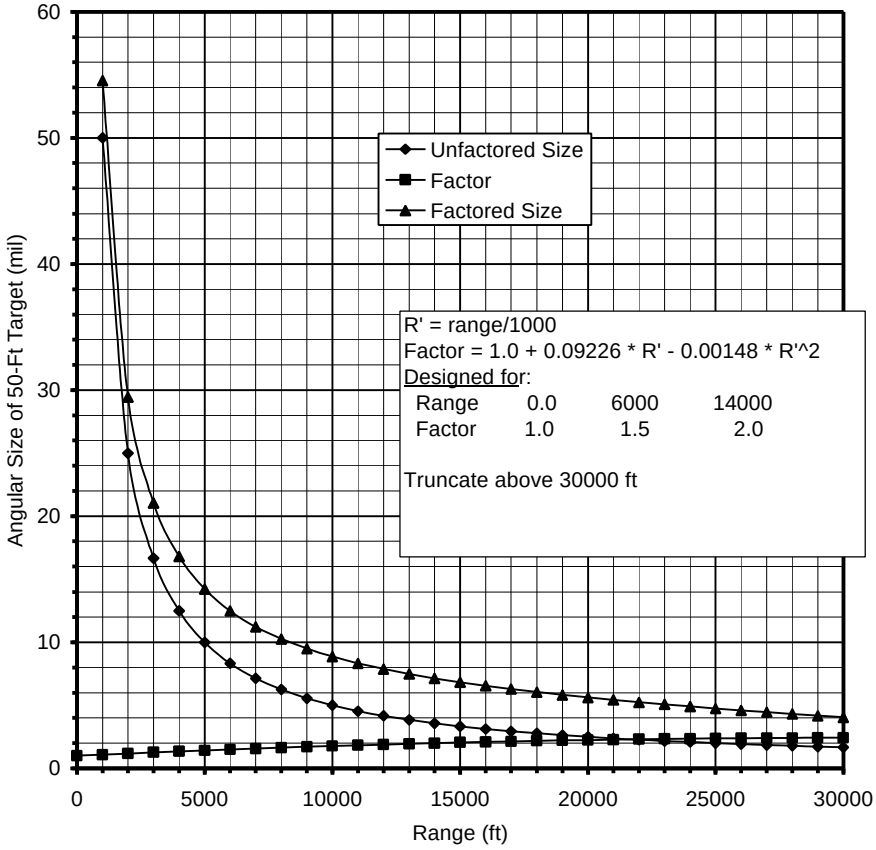

Figure 3. How aircraft size decreases with distance - factored vs. unfactored aircraft example. Factored size determined using magnification example from Figure 2.

---

[3]From D. Lerman, Warfighter Training Research Division, Air Force Laboratory, Mesa, AZ. Reprinted with permission of the author.

As mentioned at the end of Chapter II, even though a 2X or 2.5X magnification of the target seems to be an unrealistic alteration, it arguably is not (Miller, 2002). When an object is projected using only 1-2 lines of video and it's size is doubled, it still only covers 2-4 video lines, which is a barely perceivable change and almost

certainly not unrealistic in appearance. In effect, approximately only the thickness of a credit card is added to either side of the target (Miller, 2002).

A concern with this approach is that increasing the target's size might cause the pilots to wrongfully associate certain target sizes with the certain distances. This could cause a negative transfer of training. Pilots would learn to perform tactics certain ways at certain distances against an enemy that is a certain size in the simulator. However, when they went to do the same under real flying conditions, their judgements of target distance, appropriate tactics to fly and the expected outcome of their maneuvers could be in error. For example, a *real* target that appeared equal in size to a *simulated* "3-mile" target may actually only be 2 miles away. Tactics that could be used at three miles may not work at two miles considering the speeds of the two aircraft and their limited turn radius.

Of course, this is a valid concern. However, it is believed that the differences should be so perceptually small that they will have minimal, if any, negative effect. Furthermore, it is also worth reiterating that at ranges of 2-3 nmi (where the biggest size differences will be used) the simulated target is already a very rough depiction that can vary in size significantly due to aliasing. Therefore, this solution should not cause much of a change from what is already seen, and there are no known reports of negative transfer of training due to aliasing effects in this environment (Miller, 2002). Furthermore, pilots typically rely heavily on the use of a radar or a heads-up-display (HUD) to give them distance information, not solely on visual size, especially since different aircraft vary greatly in size.

A final point should be made that the reason for increasing the model's size is not because the simulator is not capable of presenting an accurate model. The problem is that

the projector lacks the lines of resolution within that target size to accurately depict the cues a pilot could normally see. Since current technology does not allow increasing the number of video lines per given area, one alternative was to increase the size of the target, thus allowing more lines of video to "paint" a given cue on the screen. However, accurate size is a problem when the target is so small that only a few lines of video are used to depict it. Due to the on/off nature of lines of video, as a very small target transitions between video lines it may be increase/decrease in apparent size by 100% or more. Targets at the sub-pixel level sometimes may be visible and other times may not, depending on where they are in their transition between lines of video.

How Will Proper Magnification Curve Be Determined?

As proposed, target size will be increased gradually as a function of distance as shown in Figure 2. The goal of this study is to determine what the real shape of this curve should be - how much should the aircraft be magnified at the various distances? To determine this, performance in the simulator needs to be matched to observed 'real world' performance, or at least to data that has been validated as a reasonable indicator of 'real world' performance. As discussed in Chapter II, existing 'real world' performance metrics can really be split into two categories. Most studies mentioned involved determining aircraft detection thresholds. The second type of performance measurement came from Warner et al. where they measured ranges for determining aircraft orientations. It is the second type of performance that is of most interest to this study. First, the user is not so concerned about detection distance since that is so variable in the real world. Secondly, due to the on/off nature of each pixel of the display, the target already illuminates a dot on the screen much longer than one would expect based on the

visibility of cues at closer ranges. In other words, it is already detectable at "appropriate" real world ranges. In fact, once a size factor model is developed to match the data from Warner et al., the aircraft may have to be decreased again in size so that it is not visible at unrealistic ranges. However, an alternative approach that prevents the aircraft from being seen too far away is the use of haze or reduction in detail that are often programmed into the simulator software. Either approach will be simple to use and the user will make that modification, if necessary, once this investigation is complete. Therefore, the primary interest is cue visibility and orientation discrimination.

To determine what aircraft size is needed for proper orientation discriminability, parts of the Warner et al. study are replicated in this study and used to provide the amount of magnification needed for distances of 0.5 to 3 miles (as covered in Warner et al). A simulated F-16 model, as nearly identical to that from the Warner et al. study as possible, is modeled and pilots' ability to determine it's orientation in the simulator is measured in much the same manner as it was done in the Warner et al. study. Four of the 16 orientations listed in Table 3 from Warner et al. are used (using more would make the experimental session too lengthy) and performance at various distances is compared to the performance listed in Table 3 for 0.5 mile to 3-mile distances. It is important to note that the target stimuli in the Warner et al. study were slides of an aircraft presented on a screen with a slide projector; they were not simulated images in a flight simulator.

Since significant time and resources are required to construct a matching simulated model, only one of the two aircraft from Warner et al. are modeled. The F-16 was chosen since performance data for it indicated a slightly wider spread in performance over the various distances. Large performance differences across distances were thought

to be desirable in helping differentiate what level of accuracy is possible at each distance. For example, the performance for one orientation seen in Table 3 may only change by 15% across all distances, while another changes over 60%. Given that experimental 'noise' is unavoidable, trying to match performance to an orientation whose performance was much different at the different distances should be much easier, and more tolerant to slight errors. For example, Orientation 4 below was diagnosed correctly only 15% of the time at 3 nmi, and 90% at 0.5 nmi. Matching performance to this large spread should be much simpler than to one that is more compressed (perhaps only a 20% performance spread across the four distances). See Figures 5-8 in Chapter V to understand how an orientation with more spread (variation along "Y" axis) should be easier and more accurate to estimate. For this reason, Orientations 4, 5, 13, and 15 (seen in Figure 4) were chosen out of the 16 possible. They showed a large and regular change in performance across the different distances. In other words, performance for the F-16 and these four orientations was found to be more *sensitive* to changes in distance than the other(s).

Subjects view these target orientations at five different distances (except Orientation 4, which needed six – explained in Chapter IV). These distances do not equate to the same ones used in Warner et al (1993). It was already known that performance in the simulator with a 'real sized' aircraft was deficient and similar performance at the same distances as Warner et al. would not be achieved. However, distances were chosen that would help produce a range of correct responses, for a given orientation, similar to those found in Warner et al. By knowing at what simulated

Figure 4. Aircraft Orientations 4, 5, 13, and 15 (From Left to Right) From Warner et al. (1993).

---

[4]From "Visual Cue Requirements for Target Orientation Assessment in Air Combat Simulation" by H. D. Warner, G. L. Serfoss, and D. C. Hubbard, 1993, AL-TR-1993-0021, Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory, pp 66-70. Reprinted with Permission.

distances performance was similar to that in Warner et al, one could determine how much of an increase in size is necessary in the simulator to achieve equal performance. For example, refer to Figure 5 in Chapter V, which shows data from Warner et al. for target Orientation 4 plotted against data from this study. To know at what distance in the simulator this target achieved performance approximately equal to what it was at 3 nmi in Warner et al., just interpolate between the data points found in this study. As Figure 5 shows, the performance from Warner et al. at 3 nmi was approximately 5.5. Tracing across the graph horizontally, a score of 5.5 would intersect a line drawn between the last two data points for this study at a distance of roughly 1.36 nmi.

Once the distance for equivalent performance is found, determining the sizing factor necessary to achieve this same performance at 3 nmi in the simulator is fairly straightforward. At 1.36 nmi, the target was 2.2 times closer and thus 2.2 times larger

than it was at 3 nmi. Therefore, to make the simulated target appear the same at 3 nmi as is does at 1.36 nmi, where desired 'real world' performance was found, it merely needs to be magnified by a factor of 2.2. It has been verified that there is no simulated haze or other condition in this simulation that would degrade the visibility of the target aircraft as a function of distance. Furthermore, the level-of-detail, which can allow the computer to present less detail on an object when it is further away was set so that the target aircraft detail was constant at all distances. Therefore, the assumption that a target twice as far away and twice its original size would appear visually identical to what it did before holds. This has been verified visually in the simulator, and is an important assumption in making the distance/size conversions mentioned above.

To get the data shown in Figures 5-8 (Chapter V), the correct/incorrect responses are taken from all 20 subjects for both times they observe that aircraft orientation at those distances. The score for each distance in Figures 5-8 thus came from 40 data points (20 subjects viewing that aircraft orientation at that distance twice). As mentioned, size magnifications needed for all four distances from Warner et al. can be determined by comparing the two curves for each orientation in Figures 5-8. Since there are four orientations being used, each of the four distances (0.5, 1, 2, 3 nmi) in Warner et al., have four size estimates attached to them (one from each orientation). The four estimates for each distance are then averaged to determine what overall size increase will be used for that distance to determine the overall magnification curve similar to the one shown in Figure 2. Therefore, the data point used for the magnification needed at three miles comes from 20 subjects observing 4 targets corresponding to that distance twice – in other words, 160 data points.

Scoring Method

Upon looking at Warner et al. one can see that performance was in the form of a dichotomous response (right or wrong). However, to gain more strength in the analysis and differentiate between slightly wrong responses and very wrong responses a continuous scoring method was developed. As can be seen in Figures 5-8, a scale of 0-7 is used. This comes from partial credit being given for each of the 3 metrics used to measure the target's orientation. For example, referring to back to Figure 1, it is very easy to mistake an aircraft with a heading of 0 degrees for one with a heading of 180 degrees. Surely the pilot mistaking one for the other still evaluated the aircraft better than one that said it had a heading of 90 degrees. Also, knowing the target was heading to the left vs. right or even 0 or 180 should receive more credit than the other responses. Therefore, for Heading, if the answer was correct, 3 of 3 points were awarded. If the pilot mistook 0 for 180, 45 for 135 or vice versa, they received a score of 2. If the target was at 90 degrees and they said 45 or 135 they received 1 point (since it is not as similar or mirror image looking as 45/135 or 0/180), and all other responses received 0 points.

For Pitch, correct answers received 2 points, and partially correct answers received 1. Partially credit was awarded for saying a pitched up target was pitched down or vice versa – at least they could tell it was pitched vs. level. Scoring Bank was similar, partial credit of 1 was given if they said a left-banked aircraft was banked right – again at least they could tell it was banked. Fully correct answered received 2 points. This scoring methodology is also represented in Table 3.

Table 3

Illustration of Scoring Methodology.

	Score			
	0	1	2	3
0 Aspect	Remaining	- -	180	0
180 Aspect	Remaining	- -	0	180
45 L Aspect	Remaining	90 L	135 L	45 L
90 L Aspect	Remaining	45/135 L	- -	90 L
135 L Aspect	Remaining	90 L	45 L	135 L
Pitch Up	Level	Pitch Down	Pitch Up	- -
Pitch Down	Level	Pitch Up	Pitch Down	- -
Pitch Level	Pitch Up/Down	- -	Level	- -
Bank Left	No Bank	Bank Right	Bank Left	- -
Bank Right	No Bank	Bank Left	Bank Right	- -
No Bank	Bank Left/Right	- -	No Bank	- -

Note. Aspect angles of 45, 90 and 135 Right are scored as a mirror image of 45, 90 and 135 Left above.

By combining a possible of 3 points for Heading, 2 for Pitch and 2 for Bank the highest possible score was 7. This allowed for previously lost information/performance differences to be captured and for a stronger analysis to be conducted. However, realize the scale is not scalar since a score of 6 is not twice as good as a 3; it is better, but not necessarily twice as good.

Chapter Summary

With data from this investigation, several magnification values for distances of 0.5, 1, 2, and 3 miles are estimated. A curve is then fit to these data similar to that in Figure 2. This indicates how much the current 'real world' sized model should be magnified at each distance to achieve orientation discrimination performance similar to actual flying conditions.

Of course, none of the 0.5 - 3 mile performance metrics mentioned are exact. They do not need to be. All that is needed is a close approximation, since, as mentioned, target visibility is highly variable even in real-world conditions. Weather, aircraft size and type, aircraft color, altitude, background, lighting and many more variables dictate that there is no constant performance metric. It is felt that the metrics used here are representative of real-world performance (based on Warner et al., 1993) and the resulting target magnification model is also a close approximation.

Chapter IV

Method

Some of the detail about which tasks were performed, why they were performed, and how they were performed was discussed in Chapter III. This chapter expands on this and describes the various elements used in the experiment and specifically how the investigation mentioned in Chapter III was conducted.

Subjects

Twenty U.S. Air Force active duty, reserve, and national guard fighter pilots participated. They were all F-16 pilots with current training in flying air combat and were also experienced in judging enemy aircraft orientations typically seen during air combat. Their ages ranged from 25 to 42. All subjects had completed a physical within the past year and were verified to have 20/20 or better vision, or were corrected to 20/20 vision with glasses. Any pilot needing correction for 20/20 vision wore his glasses for this investigation. All pilots were current on their annual air combat training requirements and a history of aircraft flown and total flight hours was gathered for each. All pilots were males. All pilots were volunteers.

Apparatus and Target

The simulator used was the A-10 Multi-Task Trainer/Display for Advanced Research and Training (DART) located at WTRD/AFRL. It used BARCO 808 projectors to rear-project an image onto frosted glass that was approximately 28 inches from the subject's eyes. Only the front simulator screen and projector were used.

The visual stimulus was an F-16 aircraft colored similarly to current F-16 aircraft in the Air Force and the F-16 model used in the Warner et al. study (1993) (see Figure 4

in previous chapter). Contrast was adjusted to attempt to match that used in Warner et al. For example, the same ratio of aircraft brightness vs. background for each orientation used was emulated in the simulator to help make the target as similar to Warner et al. as possible. Overall brightness levels of the slides could not be replicated since the projectors lacked the ability to produce the same level of brightness. Nevertheless, matching contrasts resulted in an image that looked very similar to the Warner et al. study. All 16 aircraft orientations from Warner et al. were used, but only orientations 4, 5, 13 and 15 (see Figure 4) were observed multiple times at various distances and used for comparison to Warner et al. (For a discussion of why these orientations were chosen see Chapter III). The other orientations were used merely to add variety to the task and prevent subjects from becoming too familiar with the four test orientations.

Experimental Design and Analysis

This study was set up as a single factor experiment with repeated measures. The objective was to build a descriptive model, not a predictive one. Therefore, there was no concern in building a continuous model relating the orientation discrimination capability to the range. Consequently, each orientation and range was to be analyzed separately. The data points for each orientation could then be connected as a piecewise continuous function that could be used to interpolate to as discussed in Chapter III. Since each orientation and range are to be analyzed separately, the only factor is each subject's score for the given orientation x distance combination. If this were to be a continuous model, the effect of distance would become another factor. It is a repeated measures design since more than one measurement was taken for each subject at each orientation x distance combination. This was done to maximize the amount of data gathered (for

statistical power) for a limited size subject pool. Additionally, the repeated measure provides a good estimate of measurement error discussed below along with reliability. The appropriate model for the experiment was taken from Winer (1971, p. 284) and can be expressed as:

$$X_{ij} = \pi_i + \eta_{ij}$$

where X_{ij} = observed aspect score

π_i = true aspect score

η_{ij} = error of measurement

i = each subject pilot i, $i = 1,\ldots,20$; $n = 20$

j = repetition j of a given orientation x distance combination $j = 1,2$; $k = 2$

Since each pilot saw each orientation x distance combination twice, π_i is assumed to stay constant between the two observations and η_{ij} is assumed to vary. Therefore the variance within person i is caused by measurement error, and by pooling within-person variance, one gets an estimate of variance due to error measurement (Winer, 1971). Where these measurements come from can be better illustrated in Table 4. This table will be used to illustrate the estimation of reliability for one orientation x distance combination; a similar table will exist for each of the other 21 combinations.

For the analysis of variance (ANOVA), $MS_{\text{between subjects}}$ and $MS_{\text{within subjects}}$ are defined as:

$$MS_{\text{between subjects}} = \frac{\dfrac{\sum_{i=1}^{n} P_i^2}{k} - \dfrac{G^2}{kn}}{n\text{-}1}$$

Table 4

Estimation of Reliability for One Orientation x Distance Combination

Pilot	Repetition 1	Repetition 2	Total	Mean
1	$X_{1,1}$	$X_{1,2}$	P_1	$\overline{P}_1$
2	$X_{2,1}$	$X_{2,2}$	P_2	$\overline{P}_2$
3	-	-	-	-
4	-	-	-	-
5	-	-	-	-
6	-	-	-	-
7	-	-	-	-
8	-	-	-	-
9	-	-	-	-
10	-	-	-	-
11	-	-	-	-
12	-	-	-	-
13	-	-	-	-
14	-	-	-	-
15	-	-	-	-
16	-	-	-	-
17	-	-	-	-
18	-	-	-	-
19	-	-	-	-
20	$X_{20,1}$	$X_{20,2}$	P_{20}	$\overline{P}_{20}$
Total	T_1	T_2	G	

$$MS_{\text{between subjects}} = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k} x_{ij}^{\ 2}\right) - \frac{\left(\sum_{i=1}^{n} P_i^2\right)}{k}}{(nk-1)-(n-1)(nk-1)-(n-1)}$$

Where k = number of repetitions = 2

n = number of pilots = 20

X_{ij} = each score as shown in Table 4

P_{i} = the sum of each pilot's two repetitions

G = the grand total of all scores as shown in Table 4

These measures are determined for each orientation x distance combination since they are all analyzed separately. The degrees of freedom for each is represented by the denominator of each equation above. Therefore, the degrees of freedom for $MS_{\text{between subjects}}$ is19, and the degrees of freedom for $MS_{\text{within subjects}}$ is 20 for all orientation x distance combinations. With these two MS errors determined, the estimate of variance for each orientation x distance combination is calculated the same way, using the following equation from Winer (1971):

$$\hat{s}_p^{\ 2} = \frac{MS_{\text{between subjects}} - MS_{\text{within subjects}}}{k}$$

This estimate of variance is used to produce the 2 standard deviation error bars seen on the data in Figures 5-8 in Chapter V. An example of how an ANOVA is done for an actual orientation x distance combination is presented in Table 5, Chapter V, and the accompanying text.

Reliability will also be calculated for this data. The reliability of the mean of a pilot's two observations for a given orientation x distance combination "is the variance due to true scores divided by the sum of the variance due to true scores and the variance due to the mean of the errors of measurement" (Winer, 1971). The more reliable data are, the more repeatable or consistent the results will be over time. In this case, reliability shows how consistently each subject scored for each orientation x distance combination. Did they describe it's orientation the same both times they saw it or was their evaluation different, and if so, how much different? Much of the reliability measure comes from the amount of variance between their two scores for each condition, but it also depends on comparing this value to the between subjects variance. This is illustrated in the equations below from Winer (1971, p. 287). The estimator of the reliability of the mean of each pilot's two observations is

$$r'_k = \frac{k\hat{q}'}{1 + k\hat{q}'}$$

where k is the number of measurements, which in this design is 2, and $\hat{q}'$ is defined as

$$\hat{q}' = \frac{\mathrm{MS}_{\text{between subjects}} - m\,\mathrm{MS}_{\text{within subjects}}}{km\,\mathrm{MS}_{\text{within subjects}}}$$

where $m = \dfrac{n\,(k-1)}{n\,(k-1) - 2}$

n = the number of pilots = 20

k = the number of replicates = 2

This concludes the discussion of the model for determining the variability and reliability of the data collected in the simulator. Nevertheless, there are also two other models used in this investigation. The first is the model used by Microsoft© Excel 2000 to obtain a standard deviation estimate for each orientation x distance combination in the Warner et al. data:

$$\hat{s}_{Warner} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)}{n-1}}$$

where X_i = observed aspect score

$\overline{X}$ = average score for a specific orientation x distance combination

i = subject pilot; i = 1,…,40; (n = 40)

Since there were no repeated measures, this equation for determining the standard deviation of the data is more straightforward (and likely more familiar). A standard deviation was determined from this equation for each orientation x distance combination and that value was multiplied by two. That multiplied value was then used to produce each error bar, above and below the value for each data point from Warner et al. in Figures 5-8. For example, for Orientation 4 in Figure 5, the 2 nmi data point for the Warner et al. data had a standard deviation of 0.675. Multiplied by 2, that becomes 1.35. Therefore, 1.35 is added as a vertical (score) error estimate above and below the 6.18 mean score shown (but is not shown greater than 7 since that is the maximum possible score).

The final model used is the simple one for describing a straight line

$$Y = mX + b$$

where m = slope and b = intercept.

This model was used by Microsoft© Excel to draw the straight lines between the simulator data points in Figures 5-8. Some smoothed and fitted lines were also tried with preliminary data (and also with the actual data from this study), but failed to make any appreciable difference in the resulting magnification curve. Therefore, for the sake of simplicity, and since no fit was known to be more accurate than another, straight lines were used to connect the data points and used to interpolate between the Warner et al. (1993) data as well.

As mentioned, determining where the Warner et al. scores for 0.5, 1, 2, and 3 miles distances appeared to intersect a line drawn between the simulator data points was done by tracing a horizontal line across from each of the Warner et al. scores to see where it intersected the line drawn between the gathered simulator data points. Mathematically, for each interpolation this is represented by finding the desired score on the *Y*-axis (from Warner et al.) and then using the equation for the line drawn between the appropriate two simulator data points to compute the change in *x*, when holding *y* constant at the Warner et al. value. This gives the point on the *X*-axis where performance in the simulator theoretically matches that from Warner et al.

As mentioned in Chapter III, the magnification curves are determined by visually comparing the data from the simulator to Warner et al. (1993), and graphically extrapolating between the two to determine the ratio of the distances where equal performance occurred. This ratio is used to determine how much the simulated target size should be increased. The statistical analysis mentioned here is used to determine error and reliability measures. Therefore, the statistics are not used to arrive at the actual

magnification curve shown in Figures 9 and 10. Instead, they are used to shed light on how confident one can be in that answer, based on the observed variability in the data collected here and in Warner et al.

Data Collection

Each pilot provided background flight information that included current aircraft, number of flight hours, other aircraft flown and hours for each, age, gender, and current flight status (student, instructor etc.) During the session, subjects viewed images and responded with the necessary pitch, bank, and aspect angle, which the experimenter annotated on a form. After the session, a questionnaire was used to ask pilots how realistic the images were, if they were predictable and about other experimental issues such as whether they were fatigued or distracted during the study.

Procedure

Pilots were briefed on the study and allowed to volunteer if they desired. Volunteers were asked if they wore glasses or contacts and if so, were asked to wear them during all testing. Next they filled out the background information form, and then were placed in the simulator. They were seated in a chair approximately 28 inches from the front screen, and adjusted to be at eye-height with the image. A chin rest was not used since even large variations in viewing distance appeared to have no effect on the clarity or detail of the image (it was approximately 20/40 in resolution), and there was no simple means of attaching one without a cockpit installed and without risking damage to the expensive frosted glass. Nevertheless, subjects were observed and not allowed to deviate more than a few inches from the desired eye-point location during the test.

Once subjects were seated in front of the screen, the investigator explained the task and also explained what information the pilot was to provide concerning each target. Three practice aircraft images were then presented while the investigator explained and illustrated the aspect angle, pitch and bank of each. The subject then viewed 3 additional practice targets and was asked to provide the aspect, pitch and bank of each. These practice targets were to ensure the subject understood the task completely and was clear on what information was needed and how each of the three variables was measured. Once practice was complete, the test began and each subject viewed 54 aircraft images. Each of the 4 “test” orientations (4, 5, 13 and 15) was observed twice at 5 distances: 200, 400, 800, 1600 and 2400 meters. Additionally, orientation #4 was also observed twice at 3200 meters since 200-2400 meters did not elicit the full range of performance variation seen in Warner et al. By adding the 3200-meter distance, performance dropped below that for Warner et al. at their 3 nmi distance and provided a point to extrapolate to (see Figure 5). The remaining 12 image orientations from Warner et al. were randomly assigned a range and randomly interspersed with the 42 target images. The order of all 54 aircraft images was randomized for each subject.

Each image appeared for 5 seconds and then disappeared. The subject was then given time to report the aspect angle, pitch and bank of the aircraft. Once they reported on the prior aircraft, the next aircraft was presented for 5 seconds and the process repeated. This procedure continued until all images were observed. The entire group of 54 images typically took each subject about 20 minutes or less to complete.

Once complete, subjects were asked to complete the exit survey to give their feedback about the images and the testing conditions and then were allowed to leave.

Subjects received no feedback during the testing period concerning their accuracy and were not allowed to see any images again or for longer periods.

Chapter V

Results

The scoring for each orientation and how it compares with the data gathered from Warner et al. (1993) can be seen in Figures 5-8 below. As discussed in Chapter III, the distance where simulator performance matched the 0.5, 1, 2, and 3 nmi performances from Warner et al. was taken from these graphs. That number was divided back into the appropriate distance (0.5, 1, 2, or 3 nmi) from Warner et al. to arrive at the amount of target magnification needed for that particular orientation and distance.

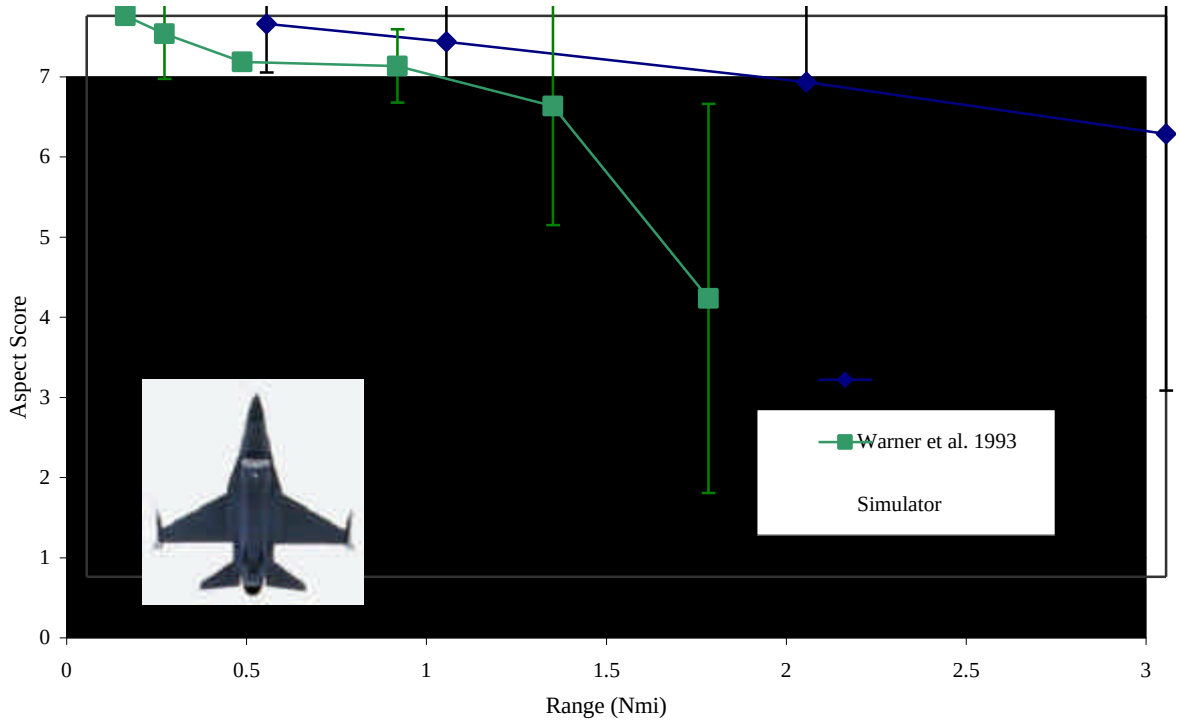

Figure 5. Performance of aircraft Orientation 4 in the simulator compared to Warner et al. (1993). Error bars represent 2 S.D. above data value and 2 S.D. below.

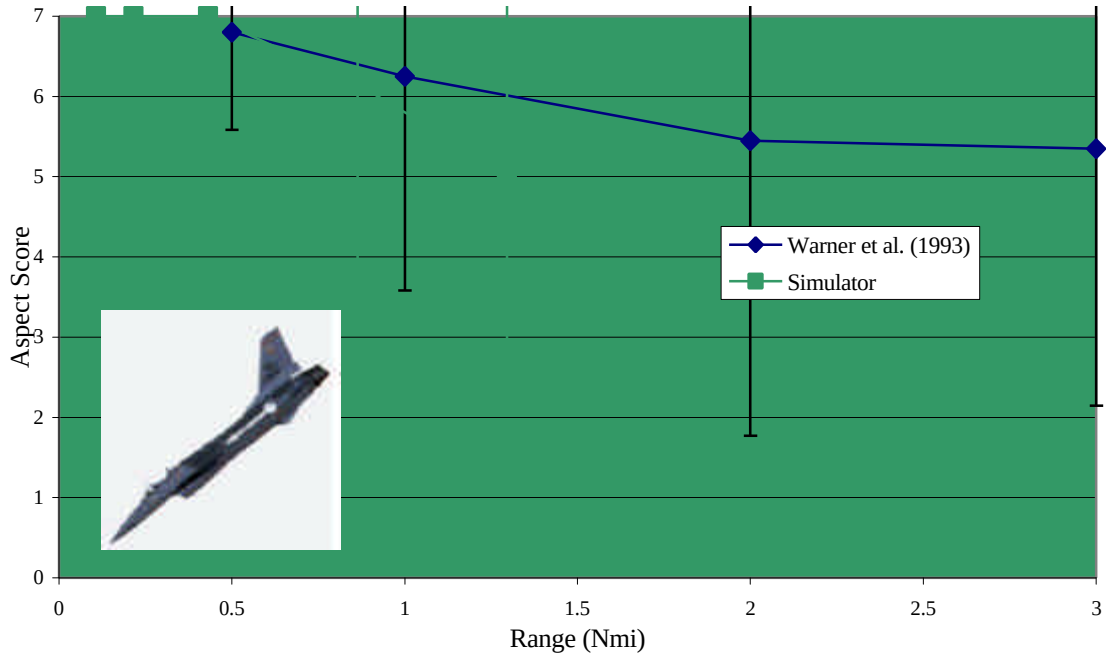

Figure 6. Performance of aircraft Orientation 5 in the simulator compared to Warner et al. (1993). Error bars represent 2 S.D. above data value and 2 S.D. below.

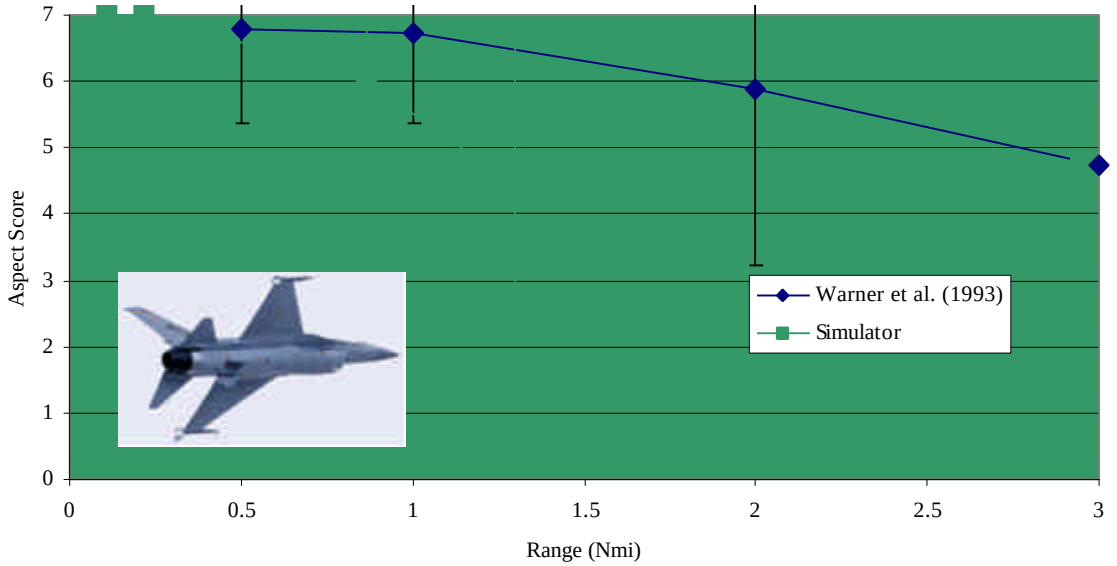

Figure 7. Performance of aircraft Orientation 13 in the simulator compared to Warner et al., (1993). Error bars represent 2 S.D. above data value and 2 S.D. below.

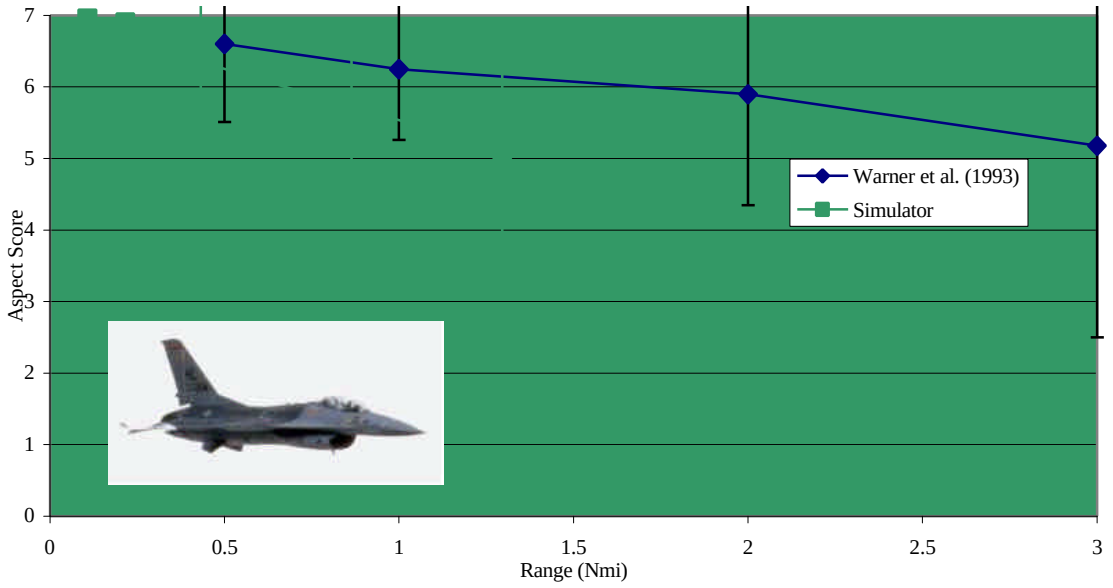

Figure 8. Performance of aircraft Orientation 15 in the simulator compared to Warner et al., (1993). Error bars represent 2 S.D. above data value and 2 S.D. below.

For clarification, another example of how this was done (in addition to the one in Chapter III), involves looking at Figure 10, which is a magnified and illustrated version of Figure 5. The score from Warner et al. (1993) for the 0.5 nmi image was 6.9, i.e. the point furthest to the left side of the graph. To determine the amount of magnification needed, a horizontal line from this point to the simulator line is drawn. The horizontal line intersects the simulator interpolated value at about 0.16 nmi. Therefore, 0.5 nmi divided by .16 nmi suggests a magnification of roughly 3.13. In other words, pilots had to be .17 nmi from the target in the simulator to equal the performance of the higher resolution slides at 0.5 nmi. Under the hypothesis of this experiment, the simulated target would thus need to be 3.13 times larger at 0.5 nmi in order for the pilot to match his

simulator detection performance to his “real world” performance. This magnification determination procedure was done for all four distances (0.5, 1, 2 and 3 nmi) for each orientation.

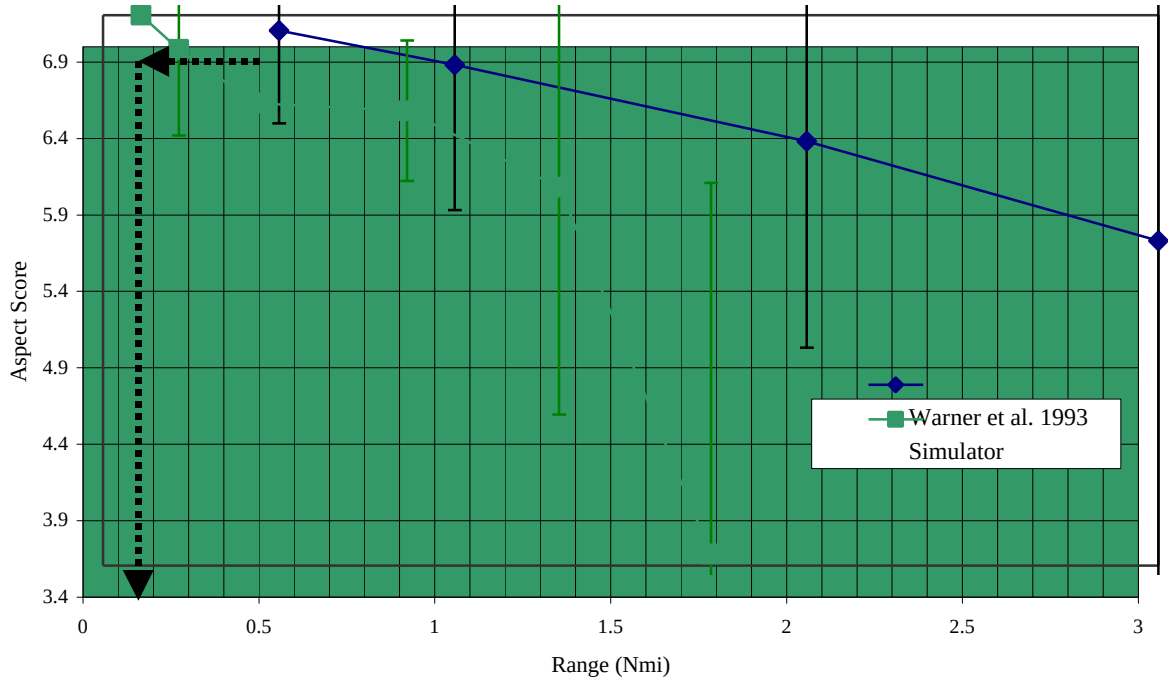

Figure 9. Performance of aircraft Orientation 4 in the simulator compared to Warner et al. (1993) with interpolation example shown.

Approximately how much each orientation would need magnified at the various distances can be seen in Figure 10. Furthermore, averaging the data from all four orientations gives a cumulative magnification line. This is represented in Figure 10 by the heavy line with the diamond shaped symbols. Figure 11 shows how this cumulative curve compares to the hypothetical one developed prior to the experiment.

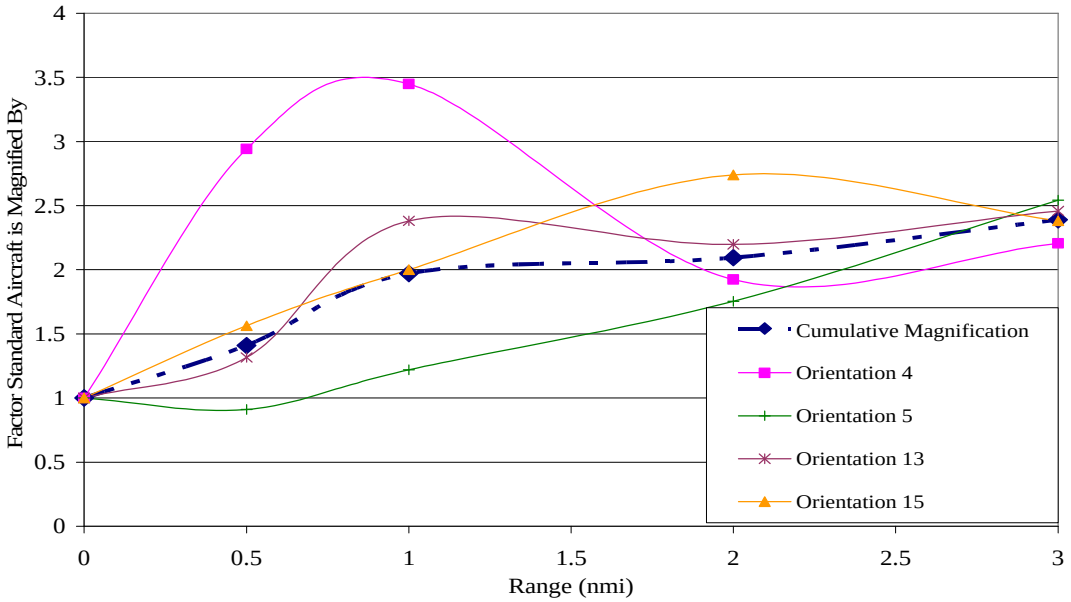

Figure 10. Amount each orientation should be magnified to match slide study performance level, and their cumulative result.

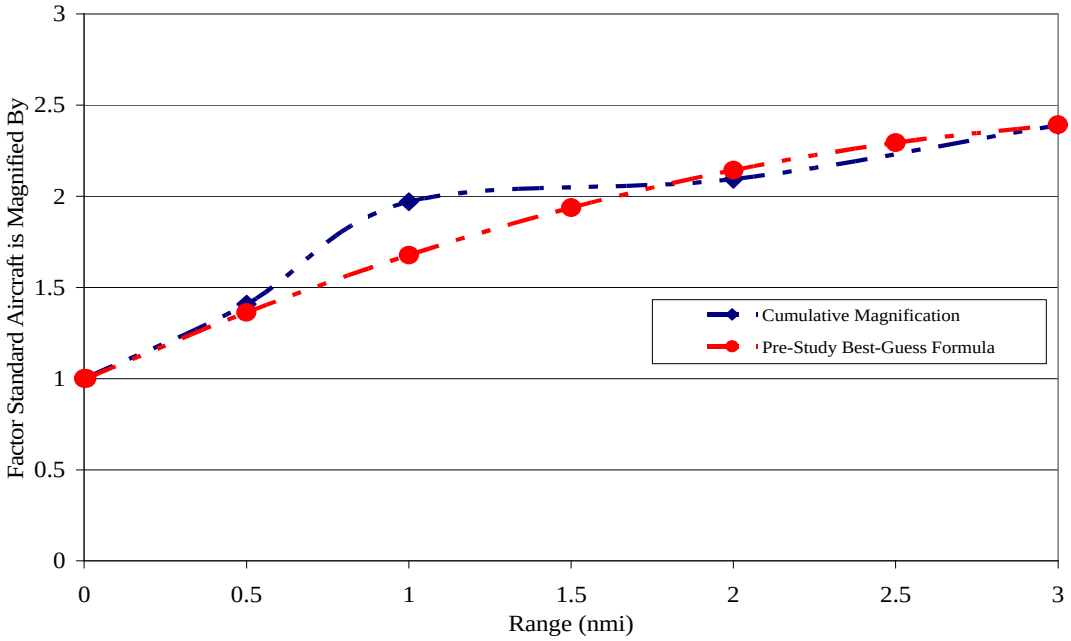

Figure 11. Cumulative magnification curve vs. pre-study “best guess” formula.

Concerning ANOVA, Table 5 presents the data for one of the orientation x distance combinations. The data in the table and the formulas presented in Chapter IV can be used to compute the variance and reliability for this data. The following is an example of how that is done:

$$\mathrm{MS}_{\text{between subjects}} = \frac{\dfrac{\sum_{i=1}^{n} P_i^2}{k} - \dfrac{G^2}{kn}}{\text{n-1}}$$

for the data in Table 5, k = 2, n = 20, $\dfrac{\sum P_i^2}{k}$ = 1632, and $\dfrac{G^2}{kn}$ = 1625.

Therefore, $\mathrm{MS}_{\text{between subjects}}$ = .368

$$\mathrm{MS}_{\text{within subjects}} = \frac{\sum(\sum X^2) - (\sum P_i^2)/k}{(nk-1)-(n-1)}$$

From Table 5, this becomes: (1637 – 1632)/20 = .25

Then, the equation to determine the variance

$$\hat{s}_p^2 = \frac{\mathrm{MS}_{\text{between subjects}} - \mathrm{MS}_{\text{within subjects}}}{k}$$

becomes: (.368-.25)/2 = 0.059 = $\hat{s}_p^2$

Of course, the standard deviation is then the square root of this number, which is 0.24. Two standard deviations thus equal 0.48, which is the size of the error bar above and below the 1600-meter data point in Figure 5.

Reliability can also be measured using the data in Table 5 and the equations in Chapter IV. Using the equation

$$m = \frac{n\,(k-1)}{n\,(k-1)-2}$$

m becomes 1.111 for all orientation x distance combinations since all have $n = 20$ subjects and $k = 2$ repetitions. The estimate of theta is determined by

$$\hat{q}' = \frac{\mathrm{MS}_{\text{between subjects}} - m\,\mathrm{MS}_{\text{within subjects}}}{km\,\mathrm{MS}_{\text{within subjects}}}$$

using the above data, this becomes: (1637 – 1.111(1632))/.556, which equals 0.1632. Reliability comes from the equation

$$r_k' = \frac{k\hat{q}'}{1 + k\hat{q}'}$$

which, using the above data gained from Table 5, becomes 2(0.1632)/(1+2(0.1632). Thus $r_k' = 0.246$ for Orientation 4 at 1600 meters. This is the reliability of the mean of the pilots' two observations of this orientation x distance combination. By applying this methodology from Winer (1971), per the above example, to all orientation x distance combinations, the results in Table 6 were obtained.

Table 5

Actual Data for Estimation of Reliability for Orientation 4 (1600 meters)

Pilot	Repetition 1	Repetition 2	Total (P_i)	Mean ($\overline{P}_i$)
1	7	6	13	6.5
2	6	7	13	6.5
3	7	6	13	6.5
4	6	6	12	6
5	6	6	12	6
6	7	6	13	6.5
7	6	6	12	6
8	7	7	14	7
9	7	7	14	7
10	6	6	12	6
11	7	6	13	6.5
12	7	7	14	7
13	5	6	11	5.5
14	7	7	14	7
15	6	7	13	6.5
16	6	7	13	6.5
17	6	6	12	6
18	6	7	13	6.5
19	6	6	12	6
20	6	6	12	6
Total	127	128	255	

Table 6

Analysis of Variance and Reliability

Orientation	Distance (meters)	$MS_{between\ subjects}$	$MS_{within\ subjects}$	s_p Standard Deviation	r_2
4	200	0	0	0	-
4	400	0.26	0.1	0.28	0.58
4	800	0.26	0.25	0	-0.1
4	1600	0.37	0.25	0.23	0.25
4	2400	1.84	0.7	0.74	0.58
4	3200	9.47	6.2	1.21	0.27
5	200	0	0	0	-
5	400	0	0	0	-
5	800	0	0	0	-
5	1600	3.68	0.9	1.17	0.73
5	2400	4.58	1.95	1.12	0.53
13	200	0	0	0	-
13	400	0	0	0	-
13	800	0.47	0.5	-	-0.2
13	1600	1.42	.95	0.46	0.26
13	2400	3.95	1.35	1.12	0.62
15	200	0.11	0.1	0	-0.1
15	400	0.11	0.1	0	-0.1
15	800	0.68	0.15	0.51	0.76
15	1600	0.68	0.35	0.4	0.43
15	2400	2.11	1.05	0.71	0.45

Chapter VI

Discussion

Overall, it appears that the goal of obtaining an empirically based magnification curve is possible and practical. More importantly, observation of this magnification algorithm once implemented into post-study air-to-air simulations, has given initial evidence that visual discrimination performance is indeed improved. Furthermore, the use of the algorithm to change the target size has not appeared to exhibit any unwanted effects or caused the target to appear unrealistic in any way. Lastly, it was somewhat satisfying to see that the final magnification curve, although only a rough estimate from data with significant variability, appears to match closely to the pre-study best-guess estimate. This apparent agreement should lend credit to both magnification curve models and more importantly, to the overall concept of using non-linear magnification to aid pilot performance. For clarification, it is worth re-iterating that although these magnification curves do magnify the target aircraft from its previous size at the various distances, the aircraft still gets smaller as it gets further away - just not as quickly as before. Although a magnification curve has been developed, it is important to discuss the limitations that apply to interpreting the resulting curve as well as discuss what other useful information can be gleaned from the data gathered.

Limitations

It is certainly recognized that this proposed magnification curve lacks fine precision, especially given the variation observed for the data points graphed in Figures 5-8. This variability limits how concretely the corresponding correcting magnifications suggested by each orientation in Figure 10 can be interpreted. Just as the error bars in

Figures 5-8 show over what area 95% of the scores are expected to fall, the magnification curves likewise contain similar variability. As seen in Figures 5-8, the variability normally tended to increase as a function of distance, as one would expect. Therefore, the magnification curves would be similarly affected. For example, the two and three mile distance estimates are more suspect and likely to vary more from what is shown than the 0.5 and 1 mile estimates. It would be more accurate to interpret the magnification curve as a shaded area that changed as a function of distance instead of as a point on a line. The area for each orientation (or the cumulative one) would likely be funnel-shaped and start off small at short distances and increase in size as distance increases. The current curve would fall within this funnel-shaped area, but the true magnification for a given distance would not be a precise point on the line, but rather a point somewhere within that funnel-shaped area for that distance. The precise size or shape of these funnels is not known and cannot be readily gleaned from Figures 5-8. In the interest of simplicity and practicality, the curved lines in Figure 10 will continued to be used since they are the still the best estimate of what the magnification should be. It will just be important to remember that each line is not precise, especially at longer ranges, and that the true value may be somewhat greater or smaller. Following are some mediating factors that discuss why great precision in this curve may not be necessary.

It was recognized at the outset of this investigation that a very precise curve would be impossible to attain. Even if it had been practical to use 200 subjects instead of 20, and the variability of the data collected in the simulator was one tenth of what it is now, the variability from the 40 subjects in Warner et al. (1993) would still be enough to prevent great precision in the magnification curve that was developed. Does this mean

the data is worthless and that the experiment was ill conceived to begin with? Certainly not. One just needs to understand the limitations of what can be taken from this data. The results lead to more of a qualitative trend of how the magnification should occur, rather than a quantitative, precise tool that could not only be used descriptively but also predictively. The alternative to no investigation was for users to press on with a magnification model similar to the "best guess" model developed before the study and shown in Figure 11 (which fortunately appears to have been an excellent approximation). The problem would have been the lack of any empirical data to help support such a model when it came under scrutiny from critics. This investigation certainly doesn't resolve the issue conclusively, but does provide a valuable qualitative estimate based on real-world data that does support the original "best guess" formula. It is a step in the right direction, and potentially enough validation for most potential users.

Another way of putting these limitations in perspective involves how users will likely implement the curve and what amount of change in the curve will be noticeable to pilots or noticeable in their performance. First, when it comes to implementation, it must be noted that even if the developed magnification curve were extremely precise with very small confidence intervals around each point, it likely would still not be implemented as is. This curve will likely only serve as a guide to potential users since other factors (besides target orientation discrimination performance) need to be considered. For example, in Figure 11, data from this study seems to suggest that targets need to be magnified more (get small less quickly) initially than the theoretical best-guess curve would require. Although this may be ideal from a discrimination performance perspective, following such a steep initial magnification curve may risk the target not

appearing to decrease in size realistically with distance (it still would decrease, just not very fast). Certainly, the best-guess curve in Figure 11 would result in an aircraft that changed size more uniformly. The implication: practitioners will likely utilize a curve somewhere between the two to ensure they get 95% of the performance, but still do not risk any unrealistic appearance or other negative side-effects. After all, flight simulation is still somewhat of an art and 95% target orientation detection performance is still a significant step ahead of what was achieved with no magnification. In other words, extreme precision in the curve is not necessary or even practically useful.

Secondly, if to match real-world conditions at two miles requires a magnification of 2.4 or 2.5 instead of 2.09, as estimated in Figure 11, would a pilot consciously notice such a change? Would their performance change significantly? Certainly at some amount of change the answer to both questions would be yes. However, just as variability in data collected prevented the determination of a precise curve, the same variability during actual simulator training makes such precision unnecessary.

Thirdly, pilots are quick to point out that dynamic (motion) cues also provide a lot of information about how the target is behaving. This does not apply to comparisons made with Warner et al. (1993), since they were also static images. However, it is not readily known how much of a target's orientation discrimination is drawn from static visual cues from the aircraft as compared to the target's motion. This is largely beyond the scope of this research, however, one interesting point should be brought up. The orientation that pilots had the most difficulty accurately describing statically in the simulator (Orientation #4) was also the one that pilots said would be most quickly cued with dynamic cues (Miller, 2003). Since it is either coming directly at the pilot or going

directly away, any ambiguity encountered from its static image would quickly be erased by how it moved relative to the observer. If heading toward the pilot, their combined airspeeds would result in an extremely quick change of target size and location. If traveling in the same direction, there would likely be very little dynamic change. The difference would be very significant and easily noticeable, more so than most other target orientations. If indeed the most difficult to view cues were the engine intake and canopy and those were most critical in discerning nose-on vs. tail-on orientations, it would be fortuitous that those same orientations benefited the most from dynamic cues. Perhaps once motion cues are added to the simulation, Orientation 4 would not need additional augmentation relative to the other three due to this phenomenon.

Finally, one must not forget the whole issue of aircraft detection and orientation discrimination is extremely variable itself. Atmospheric conditions, time of day, aircraft coloring, background and many other variables were shown in Chapter II to cause very large performance changes. This explains why some of the studies cited in Chapter II had subjects detecting aircraft only 3 nmi away while others were detecting them over 7.5 nmi away.

Certainly, in the interest of science, the goal was to develop a curve that was as precisely accurate as possible. However, within the limitations of the comparison data (Warner et al. 1993) and available subjects, great precision was not possible. Nevertheless, valuable data was obtained and, especially in the context of the mediating variables mentioned above, it does provide the desired answer with adequate precision for the task at hand.

Statistical Analysis

As already discussed, the primary concern with the data gathered is the amount of variation seen at each test condition. The method of developing the magnification curve still seems sound, as does the final result shown in Figure 11. As mentioned, discretion merely needs to be used in applying and presenting such data, since the true magnification curve could vary from what is shown. A replication of this experiment as well as Warner et al. (1993) using more subjects would be necessary to eliminate some of the uncertainly surrounding the data shown in Figures 5-10.

An explanation of some of the data in Table 6 is in order. The instances where $MS_{\text{between subjects}}$, $MS_{\text{within subjects}}$, and standard deviation (SD) had values of zero was the result of all subjects getting perfect scores of "7" both times they saw that orientation of that target, at that distance. There was no variability in these cases and therefore no reliability estimate. This is also seen for the 200 and 400-meter distances of Orientation 15, where values of 0.11 are the result of scores averaging 6.95 and 6.9 respectively for each distance – nearly everyone was scoring perfectly both times.

All of the standard deviations of 0 and 0.2 or 0.3 and below were excellent, and allowed for much more certainty in determining the proper magnification factor, although as discussed, the variation in Warner et al. (1993) still limits the precision possible. It is when the SDs were 0.7, and 1.1 and higher that there was cause to wonder where the true value would lie, where the connecting line from the previous data point should fall and at what distance the corresponding score from Warner et al. was truly achieved. This is why, for example, on Orientation 4 that it is difficult to know where a score of 5.5 (which is what Warner et al. had for the 3 nmi performance) would really be achieved, given the

variance of the last two data points collected in the simulator. The estimated magnification was 2.2, but this is truly only a rough estimate - it may also be 1.5 or 3. This is why caution in the interpretation, especially at the longer ranges, is necessary.

Another anomaly in Table 6 that should be explained is the negative (near zero) variance that was found for the 800-meter distance of Orientation 13. Table 6 shows that the $MS_{\text{within subjects}}$ was greater than the $MS_{\text{between subjects}}$. This is because nearly all subject scored perfectly except five, who scored perfectly once, but scored 5 the other time. They thought the plane was angled toward them instead of away from them and that they were seeing the top and not the bottom of the aircraft. This actually caused the data to have more variation within subjects than between. This should not be alarming. The overall average was still 6.7, with nearly all subjects scoring a perfect "7" both times. There certainly is some variability to this score, as evidenced by 5 subjects making a mistake, but the data point still seems valid and the variability should be small**.**

Looking at the reliability scores tells us more about the data. As mentioned, the more reliable data is, the more repeatable or consistent the results will be over time. The reliability measure (r^2) ranges from 0 – 1, with larger values corresponding to greater reliability. Ideally one would like to have all of the SDs as small as possible and the reliabilities as large as possible, which would imply that most of the military fighter pilot population that these results generalize to should score very close to the score that is being reported for each condition, and that all pilots should consistently score the same way each time they see a given target condition. As mentioned, the SDs are not too bad for most target orientations at the closer distances. So how reliable or consistent did the subjects appear to be? For all the near distances where SDs were zero or near zero, the

reliability should be good. The reliability equations in Chapter IV do not respond well to not having any variance in the data; they not only look for small within-subject variance, but also prefer that variance to be small compared to the between-subjects variance. When there is no variance within or between, there is no value for reliability. Nevertheless, in a practical sense, it appears that since all twenty subjects scored perfectly each time for a given orientation, one should be able to assume that pilots will consistently score the same way on that given orientation x distance combination. Therefore, it is repeatable (or reliable).

One would likely expect reliability to decrease with distance from the target since the data tended to show more variance at the greater distances. Looking at Table 6, one can see this did, indeed, hold true on many occasions, especially if one could consider the "dashed" values and –0.1 values to imply good repeatability. Nevertheless, there were also some occasions where reliability did not decrease with distance. For example, pilots had a reliability of 0.21 at 1600 meters for Orientation 4 and then at 2400 meters had a reliability of 0.56. Most of this is a reflection of between subject variation. Pilots showed almost three times more variation between their two responses at 2400 meters than they did at 1600 meters. The difference was due to the relatively small amount of variation between pilots at 1600 meters compared to the significantly greater variance at 2400 meters. Therefore, although within-subject variation increased with distance, it did not increase nearly as much as between-subject variation did. The result was a greater reliability measure at the longer distance.

Does this in any way imply that the score at 2400 meters would be more accurate or precise than the one for 1600 meters? No, the overall variance in scores for 2400

meters was still much greater than for 1600 meters. Therefore, one is less sure of exactly what score should be used at 2400 meters than at 1600 meters. For purposes of this investigation, the effect of variance seems much more important and relevant to obtaining an accurate magnification estimate than the effect of reliability.

The phenomenon of often having as much variation between a pilot's two observations of a target x distance combination as there is between different pilots' ratings of the same combination has many implications and interpretations. First, it may show how similarly all pilots (at least United States Air Force F-16 pilots) are in discrimination ability. They are all trained similarly to evaluate other aircraft and often have similar flying experience. Between-pilot variables like slight acuity differences should not have been a factor with the limited resolution of the display. This similarity in performance, of course, hurts the reliability of the data at times, but should enhance the generalizability of any findings. The fact that they all consistently score so similarly on the task means that other pilots not in the sample should be expected to perform similarly. It also shows how difficult the task can be and may also signify a threshold or chokepoint of sorts has been reached. For example, at the 1600-meter distance, nearly all subjects were scoring a 6 or 7 because they could see the aircraft well but were still perplexed about whether they were looking at the top of it or the bottom. There is little variability between subjects because they all know its pitch and bank and are debating between two aspect angles (0 and 180). They are all, more or less, guessing whether it is the top or bottom of the aircraft that they see. Many subjects, not recognizing the fact that they are looking at the exact same image, say it is coming toward them one time and going away the next time they see it. The results is within-subject variation that is almost as great as

between-subject variation. The SD is still low, but reliability suffers. Now at the 2400-meter distance, it is not just a top vs. bottom of aircraft issue. Some pilots begin to think the aircraft (which is becoming a “blob”) may be banked; others begin to wonder if it is pointed left or right instead of towards or away from them. They start to differ much more from each other with their evaluations, and although each pilot also shows more variability between their two responses for this test point, that variability is much smaller than the variability between them and the other pilots. The low reliability at 1600 meters can likely be explained as a top vs. bottom “threshold” or “chokepoint” of sorts that funneled between-subject variation very close to within-subject variation.

A final positive note concerning the similarity of between and within-subject variation is that it shows that pilots obviously were not realizing that they were seeing the same target combinations multiple times. If they did, they would have been prone to evaluate it the same way (i.e. “I’ve seen that exact one before; I’m still not sure what it is doing, but I’m going to stick with the answer I gave the first time, which is …). So from that perspective, although the reliability measurements suffer, it may have almost been disconcerting to see extremely high reliability measurements for fear that they were indeed recognizing the repetition and letting a prior response influence the latter response.

In summary, it would have been nice to see less variation in the data, especially at the greater distances, just so there would be more confidence in the accuracy of the magnification curve. Nevertheless, the goal of obtaining an empirically based magnification curve was achieved and, especially in the context of how it will be used, should have adequate precision over the full range of target distances examined. Most of

the anomalies presented in Table 6 appear to have reasonable explanations and do not point to any experimental flaws that would endanger the validity of the results. In fact, by investigating them, they merely helped bring greater understanding to the data and explain interactions and phenomenon that likely would not have been noticed otherwise. Of course, this is no surprise since this is certainly one of the greatest benefits derived from statistical analysis.

Orientation Variation

Although some of the specific numerical data analyses were just discussed, it is important to also look at the data in other ways. One of the most interesting items that begs explanation is Figure 10, with the various magnification curves for each orientations. Even when acknowledging that none of the curves are extremely precise and allowing for some variation, there still appear to be some significant differences between the various orientations. The question then, of course, is why? As seen in Figure 10, at the distances of 0.5 and 1 nmi, Orientations 4 and 5 seem to be the most different, not only from each other, but also from the group average and the pre-study "best-guess" curve.

Orientation 4. Orientation 4 needed to be magnified 3-3.5 times at the 0.5 and 1 nmi distances - much more than any other orientation. This implies that pilots had to be much closer to it than expected to make an accurate assessment of its orientation. There appear to be a few explanations for this. First, as one can see by looking at Figure 5, even though performance appears lower than that from the slide study for each distance, most of the scores, even for targets slightly over 1 nmi away, were still above 6 (out of 7). Certainly, the fact that the performance in the slide study and this one for Orientation

4 were so high causes both lines to be fairly flat. This demonstrates that performance for this orientation is not very sensitive to changes in distance, at least not initially. Therefore, even small performance differences between the simulator and the slides resulted in very large translated magnification differences. In other words, the differences between simulator performance and the slide study are probably over-exaggerated due to this lack of scoring sensitivity at shorter ranges. Furthermore, one should remember that the scoring was also somewhat crude and lacked precision. Subjects could only receive whole numbers on the scale as a score. For example, they were either a 6 or a 7, not a 6.2 or 6.6. Therefore, when a scoring difference of only 0.1 or 0.2 makes such a large difference to the magnification values, as with this orientation, and with this limited number of subjects, there likely is not enough precision in the scores to weight the results or this orientation too heavily. The fact that most of the scores are near the maximum of seven possible also implies the likelihood of a ceiling effect at work with the scoring. Finally, as stated in the limitations section above, the significant variance seen with much of this data, as well as Warner et al. (1993), also limits how in-depth or detailed any scrutiny should be. The effect of large variances in the data combined with an orientation that appears somewhat insensitive to the effect of range warrants even more caution when interpreting the data.

Nevertheless, even if some of the above arguments account for the unexpected performance of Orientation 4, one must still wonder if there is also a real effect at work and not just a measurement/sensitivity effect. What is different about Orientation 4 that would make it necessary to magnify it so much more than the other orientations, especially at the middle and closer ranges? Looking at the raw data for the shorter

distances, nearly all the scores are sixes and sevens – as expected. In the case of this orientation, almost all scores of six were simply due to mistaking the bottom of the aircraft for the top. This was typically the only error subjects were making once the aircraft was within 1 nmi, and still the primary error beyond that. They almost always correctly evaluated its pitch and bank. The question still remains, why were subjects still no better at telling the top from the bottom at 800 meters than they were at 1600 meters? Were some visual cues missing that would have helped? Some of the primary cues used to discriminate the top from the bottom of the F-16 are the presence/absence of a canopy or engine intake. Another cue is the presence of fuel tanks or items attached to the bottom of the aircraft (this model was clean (had no items) as did the aircraft in Warner et al. (1993)). A final cue (depending on paint scheme) is a color or shading difference between the top and bottom of the aircraft, with the top often being darker than the bottom. This shading difference was present on the simulated model, as it was on the aircraft in the Warner et al. slide study and many F-16s in the Air Force inventory. Nevertheless, pilots in the Warner et al. study did not appear to use color or shading differences to discern top from bottom (only 2.2% of subjects reported using that cue for evaluating Orientation 4). Although, some use of color may be nested within a stated cue like “belly” or “top of aircraft/wings” (i.e. how did they know it was belly vs. top?). Anecdotal evidence suggested that color/shading was not used much by pilots in this experiment either. Although the shading differences seemed apparent to the experimenters as they viewed the targets, many subjects failed to notice or at least failed to mention using it as a cue. A few mentioned forgetting whether the top or bottom was supposed to be darker, although this was pointed out and shown in the example slides and

coincides with how many F-16s are often painted. More often, it appeared that pilots were primarily looking for other cues to make their determinations or were not comfortable making absolute discriminations between top and bottom. For example, pilots only saw one target at a time and did not always appear comfortable deciding whether it had a “top” or “bottom” shading. The top/bottom difference was obvious in side-by-side comparisons, but when seen alone, it was not always easy for them to decide. Again, the contrast ratios of top vs. bottom shading were kept as close to Warner et al. as possible.

Since shading was similar between both studies and not apparently used, and aircraft in both studies were “clean” (without any missiles or fuel tanks attached to the bottom of the aircraft), the most obvious cue for determining top from bottom was the presence/absence of a canopy vs. engine intake. This agrees with data collected in Warner et al (1993), which showed the cue used more than twice as much as any other for evaluating Orientation 4 was the intake. The intake and canopy are very small cues compared to most of the others used: aircraft silhouette, wings, fuselage etc. The evaluation of the other orientations relies more on larger, more global aircraft cues such as aircraft silhouette, nose vs. tail, wings vs. no wings etc. The sizes of these objects are much larger than cues like the canopy and intake and thus have many more pixels that can be used to display them. This is likely why subjects had to get so much closer to the Orientation 4 model than expected. This could also explain why performance may have been no better at 800 meters than at 1600. Perhaps they saw enough of the detail of the aircraft at 1600 meters to accurately judge all other aspects of its orientation, and just were not able to see a canopy or intake. If the canopy and intake were still not

discernable at 800 meters, one would not expect better performance. It appears that subjects were not able to make out the intake or lack of canopy until about 400 meters. (This was verified using post-study observation of this target orientation and these distances, and appears to be a reasonable theory). This suggests that perhaps while a 2X magnification of the overall aircraft may be enough in many instances to provide realistic performance, when significantly smaller objects need to be detected, larger magnifications may be needed. An alternative to using a larger overall magnification would be to artificially enlarge these objects or otherwise enhance their appearance (i.e. contrasting color or shading). A simple follow-on study testing this would likely be beneficial.

Orientation 5. Figure 10 shows that Orientation 5 basically needed no magnification out to almost 1 mile for subjects to equal the performance of Warner et al (1993). The most likely explanation for this is the same one used at the end of the discussion of Orientation 4 above. Discriminating the orientation of a target is done by noticing the presence or absence or certain aircraft parts (i.e. canopy, intake, tail wings etc.) or by observing the relative shape of the object or it's relative orientation to other parts. For example, if the nose and tail are seen and the nose is lower than the tail, the aircraft is pitched down. If the wing is seen, but it doesn't appear to be as large as when viewed at a perpendicular angle the aircraft must be banked or angled away from the viewer. Orientation 4 required discerning very small aircraft parts to determine if one was looking at the top or bottom of the aircraft. Orientation 5 is basically the opposite. It requires only the largest of cues (aircraft silhouette/planform) to fully determine its orientation. It is "easy". Pilots in this study scored perfectly (7 of a possible 7 points) on

Orientation 5 at 200, 400 and 800-meter distances. Arguably, by utilizing only large, global cues, this orientation is less sensitive to decrements in resolution. It is simple to tell the nose from the tail and its slender shape and length make it obvious that no wing tops or bottoms are in view and that it is not angled toward or away from the viewer. Orientations 13 and 15 seem to fall somewhere in between 4 and 5 as far as size of cues needed and difficulty. This makes sense based on cues pilots said they used to judge these orientations in Warner et al. The likely reason there is inter-orientation variation present between this study and Warner et al. is that discrimination performance of all orientations is not affected equally when resolution is lowered.

Implications

How could performance for Orientation 4 be addressed separately? As discussed in the first few chapters, although enlarging the target (as investigated here) can hopefully create significant gains in target orientation discrimination performance, it does not need to be the sole method. As the discussion of Orientation 4 above suggests, it may be best to use color, shading, or disproportionate scaling of one particular part of the aircraft to make up for some performance limitations instead of unnecessarily enlarging the whole aircraft. For example, at the 1 nmi distance, Figure 12 shows that Orientations 5, 13 and 15 and the pre-study model suggest a magnification of 1.7. This is insufficient for the desired performance on Orientation 4, which would require nearly a 3.5 magnification (see Figure 10). Instead of compromising, and approximately doubling the target's size, as the cumulative curve in Figure 10 suggests, why not see if an additional remedy can bring the performance of Orientation 4 more in line with the other three? Perhaps the canopy and intake can be distinctly colored or shaded, or artificially enlarged. By using

remedies such as these along with the 1.7 magnification suggested by the other orientations and the model in Figure 12, one should be able to approach the performance levels found for Orientation 4 in Warner et al.

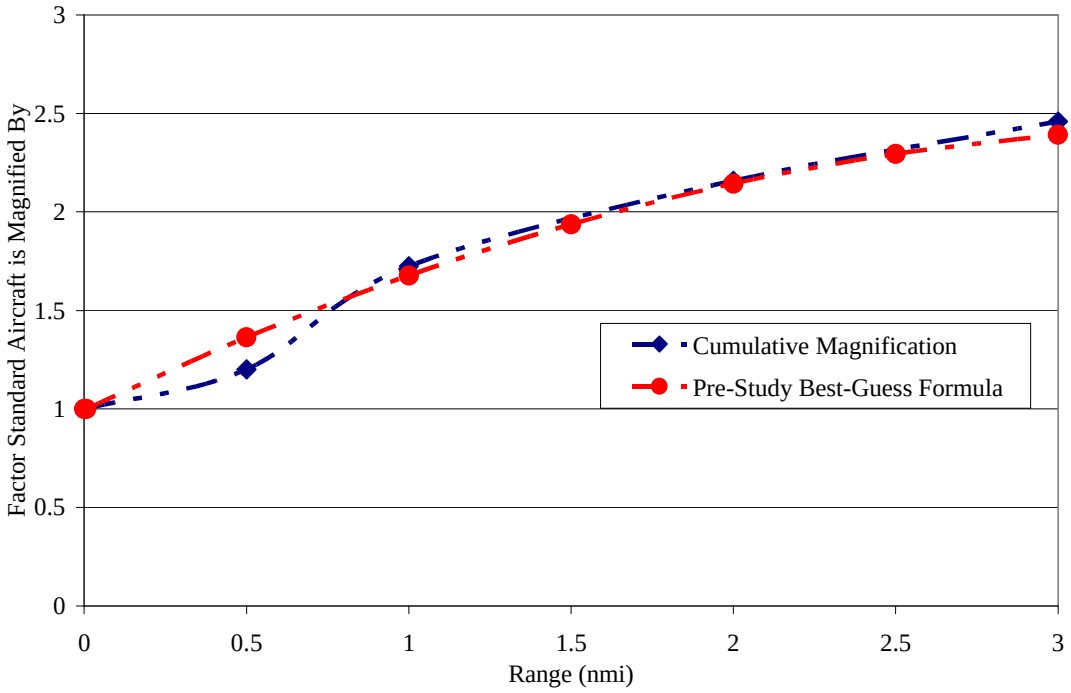

Figure 12. Cumulative magnification curve (minus orientation 4's effect) vs. pre-study "best guess" formula

It is interesting to note that Figure 12 shows data from three of the orientations matching almost perfectly with the pre-study "best guess" model. This seems to imply that if done well, size changes for other aircraft and targets/objects in general should be able to be accurately modeled without necessarily requiring data to be collected and compared to real world data for the "target" using an appropriate subject group. However, this assumes three important things. First, one has enough understanding of

the environment and “target” to develop a well-educated estimate. Secondly, special cases like Orientation 4 can to be anticipated and compensated for separately, or are could be considered insignificant enough to be ignored. Finally, due to the variability apparently inherent to this type of task, it will always be hard to know if the best-guess curve (or even an empirically derived one) exactly replicate real-world conditions.

Possible Follow-on Work

Some areas of potential research that would further shed light on this issue and this possible approach for dealing with it have been mentioned. Certainly, the observation of a target aircraft, modified as proposed here, in a dynamic air-to-air simulation will be very useful. It will help ensure that pilots do not feel the aircraft appears unrealistic in size or negatively affect the simulation in any other way. Comparing performance with a “magnified” aircraft to an “unmagnified” one would also be useful, although perhaps somewhat difficult when adding the variable of motion to the methodology.

Initially it might seem beneficial to replicate this research using a “magnified” aircraft. Certainly, any additional data would be useful, however, this approach may not yield the validation one would think. What has been determined in this investigation is the **size** the target needed to be for pilots to reach the “appropriate” level of performance. It is arguably irrelevant how that size was achieved. For example, if it was shown that pilots had to be 0.5 nmi from the target in the simulator to perform as well as pilots performed at 1 nmi in Warner et al. (1993), the result would be a recommendation to magnify the target by a factor of 2 to match the simulator performance to Warner et al. It is certainly possible to show the same simulated target at twice its’ size at 1 nmi to again

measure pilots' performance. Nevertheless, it was unnecessary to use Air Force pilots to verify that a standard sized target aircraft at 0.5 nmi looked exactly the same as a twice magnified aircraft at 1 nmi. Theoretically the aircraft size and distance may have changed, but the image requested to be shown by the projector is exactly the same. The proper screen size of the target was determined in this investigation, and as long as that screen size is repeated, the performance should stay the same, regardless of the mathematical calculations of size and distance used to achieve it.

Certainly, it may be beneficial to collect data closer to the suspected points where the Warner et al. data horizontally intersect the simulator data lines in Figures 5-8. This would help eliminate some of the possible error that may have been made when interpolating between existing data points. Nevertheless, given the variation seen in this data as well as the Warner et al data, further refinement would be difficult until much more data could be collected and the variation in each reduced.

One other area of potential interest was to replicate this investigation but place the pilots further (perhaps twice as far) away from the screen and make all the images larger (i.e. twice as big). The effect should be a target that still subtends the same visual angle (looks the same size), but that would have roughly twice as many pixels present to project the image. The resolution would be much greater and its effect on performance could be better evaluated.

Of course there are many other avenues and variables that could be pursued. These were just a few of the ones that seemed to be appropriate things to look into.

Chapter VII

Conclusions and Recommendations

This is certainly an application oriented approach and result, more so than theoretical. Lacking a better solution or another obvious approach to solving the problem of inadequate resolution for air-to-air target orientation discrimination, this method appears to have achieved what was desired. As discussed in Chapter II, the military aviation air combat environment is highly variable and so is visual detection/discrimination performance. It is believed that the use of this magnification model, in the context of the resolution limits of today's image projectors, provides a more reasonable "simulation" of flying against a real-world enemy, and will now allow better air combat training for U.S. military pilots.

In a more general sense, this study should demonstrate that it can be effective (and apparently nearly imperceptible based on initial anecdotal evidence), to have an object change size in a non-linear fashion as a function of distance. This method can be a means of overcoming inadequate resolution to portray images more realistically at real-world distances. Certainly, such a method could also be used to "magnify" simulated ground images or other targets that a pilot needs to observe or "target" from long distances. Furthermore, this should reach beyond the realms of flight simulation to other types of simulation, whether it is a simulated medical surgery or a consumer video game. When the resolution of a system cannot provide the necessary detail for an object of a given size, it appears that, up to a point, size can be enhanced to compensate for this, yet the realism of the simulation need not be negatively affected. There are certainly limits to how much size enhancement can occur and still appear realistic. Furthermore,

although this method may be the first, and best method (other than better displays), to deal with this issue, other target enhancements such as color, shading or other forms of cueing may also need to be considered to enhance overall performance and to prevent possible side-effects of too much target magnification. Finally, if other applications were to be tested in this manner they would certainly benefit from the availability of real-world data, such as Warner et al. (1993), to work toward matching, and to also help validate the developed model.

Does the potential application of this technique disappear as technology erases the display resolution hurdle? Arguably, the answer is no, at least for the foreseeable future. As discussed earlier, even when newer technology does eliminate the need for such a "work-around" solution, the cost of such technology can often still be prohibitive. Justifying this expense is especially difficult in instances where the current hardware is capable of accomplishing the majority of the necessary tasks. There may be no need to replace a good, and otherwise competent, system just to accomplish a few "visually difficult" tasks. Furthermore, whether it is flying-squadron-level "low-fidelity" flight trainers or a scaled-down portable version of a video game, there has always been the market for lower cost, more portable versions of otherwise expensive simulations. Therefore, until even the least expensive, most portable displays can provide more detail than the human eye needs, this type of visual performance-enhancing method should be useful.

References

Bloomfield, J. R., Smith, D. A. (1982). *Evaluation of Training Techniques for Air-to-Air Target Acquisition* (SAM-TR-82-32). Brooks AFB, TX: Aerospace Medical Division, USAF School of Aerospace Medicine.

Boff, K. R., Lincoln, J. E. (1988). *Engineering Data Compendium: Human Perception and Performance, Volume II*. Wright Patterson AFB, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Buffett, A. R. (1986). Visual Cuing Requirements in Flight Simulation. In: Proceedings of the *International Conference on Advances in Flight Simulation – Visual and Motion Systems*. London, England: Royal Aeronautical Society, 127-157.

Costanza, E. B., Stacey, S. R., Snyder, H. L. (1980). *Air-to-Air Target Acquisition: Factors and Means of Improvement* (SAAM-TR-80-9, AD-A087 848). Brooks AFB, TX: Aerospace Medical Division, USAF School of Aerospace Medicine.

Coward, R. E., Rupp, A. M. (1982). *Simulators for Air-to-Air Combat Versus Real World: Visual Cue Analysis for Simulated Air-to-Air Combat Training* (AFHRL-TR-81-26, AD-A110 570). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Hamilton, P. V., Monaco, W. A. (1986). Improving air-to-air target detection. *Wings of Gold*, 46-48.

Hoffman, H. E. (1976). *A Review of the Most Important Established Facts About the Visibility (Maximum Detection Range) of Aircraft* (Library Translation No 1895). Great Britain: Royal Aircraft Establishment.

Hutchins, Jr., C. W. (1978). *The Relationship between Air Combat Maneuvering Range (ACMR) Output Measures and Initial Visual Acquisition Performance* (NAMRL-SR-79-1, AD-A062 134). Naval Air Station, Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Kress, G., Brictson, C. A. (1983). *Operational Analysis of Visual Skills for Air Combat Maneuvering* (NAVTRAEQUIPCEN 80-D-0011/041-3). Orlando, FL: Code N-712, Naval Training Equipment Center.

Miller, M. (2002). Personal Interview. Air Force Pilot and United States Air Force Reserve Simulation Director, Warfighter Training Research Division, Air Force Research Laboratory, Mesa, AZ.

Miller, M. (2003). Personal Interview. Air Force Pilot and United States Air Force Reserve Simulation Director, Warfighter Training Research Division, Air Force Research Laboratory, Mesa, AZ.

Monaco, W. A., Hamilton, P. V. (1985). Visual Capabilities Related to Fighter Aircraft Performance in the F-14 and Adversary Aircraft (AGARD-CP-395). In: *Medical Selection and Physiological Training of Future Fighter Aircrew, AGARD Conference Proceedings*. Neuilly Sur Seine, France: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development)

O'Neal, M. R., Miller, II, R. E. (1998). *Further Investigation of Contrast Sensitivity and Visual Acuity in Pilot Detection of Aircraft* (AAMRL-TR-88-002, AD-A198 434). Wright Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Pierce, B. (2002). Personal Interview. Visual Research Director, Warfighter Training Research Division, Air Force Research Laboratory, Mesa, AZ.

Provines, W. F., Rahe, A. J., Block, M. G., Pena, T., Tredici, T. J. (1983). *Yellow Ophthalmic Filters in the Visual Acquisition of Aircraft* (USAFSAM-TR-83-46, AD-A138 536). Brooks AFB, TX: Aerospace Medical Division, USAF School of Aerospace Medicine.

Temme, L. A., Still, D. L. (1991). Prescriptive Eyeglass Use by U.S. Navy Jet Pilots: Effects on Air-to-Air Target Detection. *Aviation, Space, and Environmental Medicine*, *62*, 823-826.

Warner, H. D., Serfoss, G. L., Hubbard, D. C. (1993). *Visual Cue Requirements for Target Orientation Assessment in Air Combat Simulation*, (AL-TR-1993-0021). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Whiner, B. J. (1971). *Statistical Principles in Experimental Design*, (2nd ed., pp. 283-289). New York: McGraw-Hill.